

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-099722

(43)Date of publication of application : 07.04.2000

(51)Int.Cl.

G06T 7/00

(21)Application number : 10-268419

(71)Applicant : TOSHIBA CORP

(22)Date of filing : 22.09.1998

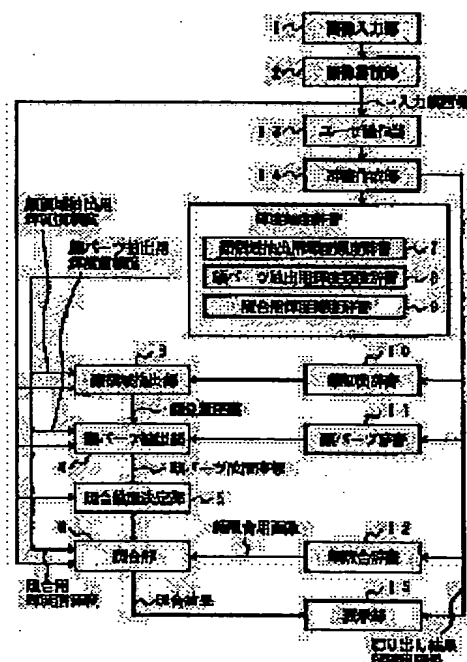
(72)Inventor : ENOMOTO NOBUYOSHI

## (54) PERSONAL FACE RECOGNIZING DEVICE AND ITS METHOD

### (57)Abstract:

**PROBLEM TO BE SOLVED:** To make it possible to execute quick and accurate recognition processing for the image of a personal face.

**SOLUTION:** The image recognition device is provided with an fetching part 1 for fetching a target image, a reliability calculation part for comparing dictionary data 7 to 9 related to previously stored luminance frequency with the luminance values of plural unit areas of the target image and obtaining the reliability of plural unit areas of the target image, and a collation part 6 for comparing a face collation dictionary 12 previously stored for recognition with the target image in every unit area based on the calculated reliability and collating the target image.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

**\* NOTICES \***

**Japan Patent Office is not responsible for any damages caused by the use of this translation.**

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

---

**CLAIMS**

---

[Claim(s)]

[Claim 1] Image recognition equipment characterized by providing the following. An image input means to input an image for recognition in image recognition equipment which recognizes an image for recognition. A dictionary creation means to draw up a dictionary for collating from an image for recognition inputted with this input means. A brightness frequency dictionary held in quest of brightness frequency of each pixel in an image for recognition based on two or more dictionary data created with this creation means. \*\* which has a collating means which collates by considering reliability about each pixel of an input image based on brightness of each pixel of said brightness frequency dictionary and input image in the case of this collating while collating an input image and a dictionary drawn up with said creation means at the time of recognition of an image for recognition.

[Claim 2] It is image recognition equipment characterized by making it not use said collating means for collating in image recognition equipment according to claim 1 about a pixel which seemingly will not be reasonable as an image for recognition based on said brightness frequency dictionary.

[Claim 3] It is image recognition equipment characterized by said collating means making a contribution to a collating result low in image recognition equipment according to claim 1 about a pixel which seemingly will not be reasonable as an image for recognition based on said brightness frequency dictionary.

[Claim 4] It is image-recognition equipment which said brightness frequency dictionary has a brightness frequency dictionary for a face field extract held in quest of brightness frequency of a logging image at the time of dictionary data origination by said creation means in image-recognition equipment according to claim 1, and is characterized by for said extract means to have a face field extract means which cuts down an image for recognition based on said brightness frequency dictionary.

[Claim 5] In image recognition equipment according to claim 4 said dictionary creation means A parts dictionary about a parts image of a specific region in an image for recognition is drawn up. Said said brightness frequency dictionary has a brightness frequency dictionary for a kind parts extract held in quest of brightness frequency of a parts image of parts dictionary creation time by said creation means. Said extract means is image recognition equipment characterized by having a kind parts extract means which cuts down an image for recognition to a kind parts image extracted with said face field extract means based on said brightness frequency dictionary.

[Claim 6] In image recognition equipment according to claim 5, said dictionary creation means draws up a face collating dictionary about an image for recognition of a specific region in an image for recognition. Said said brightness frequency dictionary has a brightness frequency dictionary for face collating held in quest of brightness frequency of an image for recognition of a face collating dictionary by said creation means. Said extract means has a collating positioning means which chooses an image for recognition of a specific region in an input image from relation with a kind parts image extracted with a kind parts extract means as an image for collating. Said collating means is image recognition equipment characterized by being based based on said brightness frequency dictionary for face collating, and collating using a pixel which will seemingly be reasonable as an image for recognition when collating a face collating dictionary and an image for collating.

[Claim 7] In an image recognition method of recognizing an image for recognition, a dictionary for

collating is drawn up from an inputted image for recognition. While holding in quest of brightness frequency of each pixel in an image for recognition based on two or more of these created dictionary data and collating an input image and said dictionary at the time of recognition of an image for recognition An image recognition method characterized by collating by considering reliability about each \*\*\*\* of an input image based on brightness of each pixel of said brightness frequency dictionary and input image on the occasion of this collating.

[Claim 8] It is image recognition equipment characterized by making it not use for collating in an image recognition method according to claim 7 about a pixel which seemingly will not be reasonable as an image for recognition based on said brightness frequency dictionary when collating said input image and said dictionary.

[Claim 9] It is image recognition equipment characterized by making a contribution to a collating result low in an image recognition method according to claim 7 about a pixel which seemingly will not be reasonable as an image for recognition based on said brightness frequency dictionary when collating said input image and said dictionary.

[Claim 10] In an image recognition method of recognizing an image for recognition, an image field for recognition is started from an inputted image. A face collating dictionary for collating is drawn up from an image for recognition of a specific region in an image of a started image field for recognition. In quest of brightness frequency of each pixel in a \*\*\*\*\* (ed) image field for recognition, a brightness frequency dictionary for a face field extract is drawn up. At the time of recognition of an image for recognition, a field which will seemingly be reasonable as an image for recognition is extracted as an image for recognition based on said brightness frequency among images in an input image. An image recognition method characterized by choosing an image for collating from an image for recognition of a specific region in this extracted image for recognition, and making it collate with said face collating dictionary.

[Claim 11] In case said face collating dictionary is drawn up, while drawing up a parts dictionary about a parts image of a specific region in an image for recognition in image recognition equipment according to claim 10 In drawing up a brightness frequency dictionary for a parts extract in quest of brightness frequency of a parts image of parts dictionary creation time and performing collating with said face collating dictionary and input image Based on a brightness frequency dictionary for a parts extract, and a parts dictionary, a parts image of an input image to a specific region is extracted from an image for recognition extracted based on said brightness frequency dictionary for a face field extract. An image recognition method characterized by choosing physical relationship of an extracted parts image to an image for collating from an input image, and making it collate with said face collating dictionary.

[Claim 12] The image-recognition method characterized by to collate based on said brightness frequency dictionary for face collating using the pixel which will seemingly be reasonable as an image for recognition when \*\*\*\*\* the brightness frequency dictionary for face collating held in an image-recognition method according to claim 10 in quest of brightness frequency of an image for recognition of a face collating dictionary in case said face collating dictionary is drawn up and collating said face collating dictionary and the image for collating.

[Claim 13] Image recognition equipment which is characterized by providing the following and which recognizes an image for recognition An image input means to input an image for recognition A focus extract means to extract the focus from a sample image inputted with this input means A dictionary creation means to form a partial dictionary pattern showing characteristic quantity of a partial image near [ which was extracted with this focus extract means / each ] the focus, and to draw up a partial dictionary of each focus A collating means to perform collating with characteristic quantity which extracted each focus from an input image inputted with said input means with said focus extract means at the time of recognition of an image for recognition, and was obtained from a partial image near [ each ] the focus, and a partial dictionary pattern of said partial dictionary

[Claim 14] A geometry change of a point with a unique brightness [ set to image recognition equipment of claim 13, and / means / said / focus extract ] change in a small field in an image for recognition or brightness is image recognition equipment characterized by extracting a unique point as the focus.

[Claim 15] Image recognition equipment characterized by holding only a partial dictionary pattern

which evaluated redundancy about two or more partial dictionary patterns obtained from a sample image, and was made into non-redundancy in image recognition equipment of claim 13 at the time of dictionary construction by said dictionary creation means in a partial dictionary.

[Claim 16] It is image recognition equipment characterized by complementing characteristic quantity in image recognition equipment of claim 13 using an approximation function which does not hold two or more partial dictionary pattern \*\*\*\* obtained from a sample image, but holds a partial dictionary pattern in which characteristic quantity is shown for every predetermined gap, and is complemented about the section in the meantime at the time of dictionary construction by said dictionary creation means.

[Claim 17] It is image recognition equipment which creates and holds a contribution coefficient which contributes to recognition according to a use to each partial dictionary pattern in image recognition equipment of claim 13, and is characterized by said collating means collating by considering a contribution coefficient.

[Claim 18] Image recognition equipment characterized by to use a set of a partial dictionary pattern effective in recognition on the occasion of collating by said collating means, creating and holding a contribution coefficient which contributes to recognition according to a use to each partial dictionary pattern in image recognition equipment of claim 13, and choosing from distribution of a contribution coefficient according to a use of recognition.

[Claim 19] While having a brightness frequency dictionary held in quest of brightness frequency of each pixel in an image for recognition based on dictionary data in image recognition equipment of claim 13 and collating an input image and a dictionary drawn up with said creation means at the time of recognition of an image Image recognition equipment characterized by having having a collating means which collates by considering reliability about each pixel of an input image based on brightness of each pixel of said brightness frequency dictionary and input image in the case of this collating.

[Claim 20] In an image recognition method of recognizing an inputted image for recognition, the focus is extracted from an inputted sample image. Form a partial dictionary pattern showing characteristic quantity of a partial image near [ this / that was extracted / each ] the focus, and a partial dictionary of each focus is drawn up. An image recognition method characterized by performing collating with characteristic quantity which extracted said focus from an inputted input image at the time of recognition of an image for recognition, and was obtained from a partial image near [ each ] the focus, and a partial dictionary pattern of said partial dictionary.

[Claim 21] An image recognition method characterized by setting to said focus extract and extracting a point with a unique brightness change in a small field in an image for recognition, or a point with a unique geometry change of brightness as the focus in an image recognition method of claim 20.

[Claim 22] An image recognition method characterized by holding only a partial dictionary pattern which evaluated redundancy about two or more partial dictionary patterns obtained from a sample image by said dictionary creation time, and was made into non-redundancy in an image recognition method of claim 20 in a partial dictionary.

[Claim 23] It is the image recognition method characterized by complementing characteristic quantity using an approximation function which does not hold two or more partial dictionary pattern \*\*\*\* obtained from a sample image by said dictionary creation time, but holds a partial dictionary pattern in which characteristic quantity is shown for every predetermined gap, and is complemented about the section in the meantime in an image recognition method of claim 20.

[Claim 24] It is the image recognition method which creates and holds a contribution coefficient which contributes to recognition according to a use to each partial dictionary pattern in an image recognition method of claim 20, and is characterized by said collating means collating by considering a contribution coefficient.

[Claim 25] An image recognition method characterized by to use a set of a partial dictionary pattern effective in recognition on the occasion of collating by said collating means, creating and holding a contribution coefficient which contributes to recognition according to a use to each partial dictionary pattern in an image recognition method of claim 20, and choosing from distribution of a contribution coefficient according to a use of recognition.

[Claim 26] The image-recognition method characterized by to have collating by considering reliability about each pixel of an input image in the case of this collating based on the brightness of each pixel of said brightness frequency dictionary and input image while draw up the brightness frequency dictionary which holds in quest of brightness frequency of each pixel in an image for recognition based on dictionary data in the image-recognition method of claim 20 and collating an input image and said dictionary at the time of recognition of an image.

[Claim 27] An image recognition method characterized by providing the following A taking-in production process which captures an image of an object containing a person's face image A focus detection production process of detecting the focus about an object image captured at the aforementioned taking-in production process based on the singular point of distribution of brightness in a predetermined field of this object image A contribution coefficient production process of computing two or more contribution coefficient which shows effectiveness for collating about two or more set semi- data of the focus for recognition stored beforehand, respectively, and storing it The collating production process which compares the remaining two or more set semi-data at the time of redundant criteria data being deleted at said criteria data deletion production process of deleting two or more redundant criteria data in consideration of the redundancy of criteria data, and said criteria data deletion production process with the focus detected at said focus detection production process for two or more of said unit fields of every based on said contribution coefficient stored at said contribution coefficient production process, and collates said object image

---

[Translation done.]

**\* NOTICES \***

**Japan Patent Office is not responsible for any damages caused by the use of this translation.**

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

---

**DETAILED DESCRIPTION**

---

**[Detailed Description of the Invention]**

**[0001]**

**[The technical field to which invention belongs]** This invention relates to the person face recognition equipment which is image recognition equipment which recognizes the target image pattern, and recognizes especially human being's face, and its person face recognition method.

**[0002]**

**[Description of the Prior Art]** recently, although a user has the needs of little individual collating of a burden by non-contact in the field which went as the password of OA equipment, such as individual collating of social systems, such as close recession management, an ATM device, and a ticket machine, and a personal computer, the image recognition \*\*\*\* method attracts attention as what does not use not much special equipment in that case.

**[0003]** As for this, what prepares two or more face images of a test subject, considers as a logging dictionary, performs pattern matching of an input face image and a dictionary at the time of recognition, and performs face location extract and collating is common. As a dictionary, in order to make it correspond to various conditions in each people, there are some which prepare the image of two or more sheets, but since dictionary size becomes huge if it remains as it is, there is a method which draws up the dictionary which carried out the feature compression of these statistically, and what is similar to analysis of a principal component as statistical compression is used.

**[0004]**

**[Problem(s) to be Solved by the Invention]** However, by the method by the above-mentioned conventional image recognition, there is a problem that exact recognition is difficult, in the environment which does not have little lighting fluctuation. Lighting fluctuation has brightness change in the whole screen, and the partial change by brightness inclination or the shadow, and the former is received. Although there is an example which performed normalization by the image magnitude of a vector (norm) about the dictionary and the input image, or performed histogram flattening and there is an example which carried out the plane reliance panel of the brightness in a screen to the latter There is a problem that a face extract goes wrong even if it carries out with them, or collating becomes incorrectness.

**[0005]** By being made that the above-mentioned problem should be solved, calculating the reliability for every smallness field of the field which matches, and performing matching according to this reliability, this invention realizes quick and exact matching and aims at offering the person face recognition equipment which performs recognition processing quick [ a person's face ] and exact by this, and this person face recognition method.

**[0006]**

**[Means for Solving the Problem]** An image input means by which this invention inputs an image for recognition in image recognition equipment which recognizes an image for recognition, A dictionary creation means to draw up a dictionary for collating from an image for recognition inputted with this input means, At the time of a brightness frequency dictionary held in quest of brightness frequency of each pixel in an image for recognition based on two or more dictionary data created with this creation means, and recognition of an image for recognition, while collating an input image and a dictionary drawn up with said creation means It is image recognition equipment characterized by having having a collating means which collates by considering reliability about each pixel of an input

image based on brightness of each pixel of said brightness frequency dictionary and input image in the case of this collating.

[0007] Moreover, this invention is image recognition equipment characterized by making it not use said collating means for collating about a pixel which seemingly will not be reasonable as an image for recognition based on said brightness frequency dictionary in image recognition equipment according to claim 1.

[0008] Moreover, this invention is image recognition equipment characterized by said collating means making a contribution to a collating result low about a pixel which seemingly will not be reasonable as an image for recognition based on said brightness frequency dictionary in image recognition equipment according to claim 1.

[0009] Moreover, it is image-recognition equipment characterized by for this invention to have a brightness frequency dictionary for a face field extract which holds said brightness frequency dictionary in quest of brightness frequency of a logging image at the time of dictionary data origination by said creation means in image-recognition equipment according to claim 1, and for said extract means to have a face field extract means which cuts down an image for recognition based on said brightness frequency dictionary.

[0010] This invention is set to image recognition equipment according to claim 4. Moreover, said dictionary creation means A parts dictionary about a parts image of a specific region in an image for recognition is drawn up. Said said brightness frequency dictionary has a brightness frequency dictionary for a kind parts extract held in quest of brightness frequency of a parts image of parts dictionary creation time by said creation means. Said extract means is image recognition equipment characterized by having a kind parts extract means which cuts down an image for recognition to a kind parts image extracted with said face field extract means based on said brightness frequency dictionary.

[0011] Moreover, this invention draws up a face collating dictionary concerning [ said dictionary creation means ] an image for recognition of a specific region in an image for recognition in image recognition equipment according to claim 5. Said said brightness frequency dictionary has a brightness frequency dictionary for face collating held in quest of brightness frequency of an image for recognition of a face collating dictionary by said creation means. Said extract means has a collating positioning means which chooses an image for recognition of a specific region in an input image from relation with a kind parts image extracted with a kind parts extract means as an image for collating. In case said collating means collates a face collating dictionary and an image for collating, it is image recognition equipment characterized by being based based on said brightness frequency dictionary for face collating, and collating using a pixel which will seemingly be reasonable as an image for recognition.

[0012] Moreover, this invention draws up a dictionary for collating from an inputted image for recognition in an image recognition method of recognizing an image for recognition. While holding in quest of brightness frequency of each pixel in an image for recognition based on two or more of these created dictionary data and collating an input image and said dictionary at the time of recognition of an image for recognition It is the image recognition method characterized by collating by considering reliability about each \*\*\*\* of an input image based on brightness of each pixel of said brightness frequency dictionary and input image on the occasion of this collating.

[0013] Moreover, in an image recognition method according to claim 7, in case this invention collates said input image and said dictionary, it is image recognition equipment characterized by making it not use for collating about a pixel which seemingly will not be reasonable as an image for recognition based on said brightness frequency dictionary.

[0014] Moreover, in an image recognition method according to claim 7, in case this invention collates said input image and said dictionary, it is image recognition equipment characterized by making a contribution to a collating result low about a pixel which seemingly will not be reasonable as an image for recognition based on said brightness frequency dictionary.

[0015] Moreover, this invention starts an image field for recognition from an inputted image in an image recognition method of recognizing an image for recognition. A face collating dictionary for collating is drawn up from an image for recognition of a specific region in an image of a started image field for recognition. In quest of brightness frequency of each pixel in a \*\*\*\*\* (ed) image

field for recognition, a brightness frequency dictionary for a face field extract is drawn up. At the time of recognition of an image for recognition, a field which will seemingly be reasonable as an image for recognition is extracted as an image for recognition based on said brightness frequency among images in an input image. It is the image recognition method characterized by choosing an image for collating from an image for recognition of a specific region in this extracted image for recognition, and making it collate with said face collating dictionary.

[0016] Moreover, in case this invention draws up said face collating dictionary, while it draws up a parts dictionary about a parts image of a specific region in an image for recognition in image recognition equipment according to claim 10 In drawing up a brightness frequency dictionary for a parts extract in quest of brightness frequency of a parts image of parts dictionary creation time and performing collating with said face collating dictionary and input image Based on a brightness frequency dictionary for a parts extract, and a parts dictionary, a parts image of an input image to a specific region is extracted from an image for recognition extracted based on said brightness frequency dictionary for a face field extract. It is the image recognition method characterized by choosing physical relationship of an extracted parts image to an image for collating from an input image, and making it collate with said face collating dictionary.

[0017] Moreover, in an image recognition method according to claim 10, in case this invention draws up said face collating dictionary A brightness frequency dictionary for face collating held in quest of brightness frequency of an image for recognition of a face collating dictionary is \*\*\*\*\* (ed). In case said face collating dictionary and an image for collating are collated, it is the image recognition method characterized by collating using a pixel which will seemingly be reasonable as an image for recognition based on said brightness frequency dictionary for face collating.

[0018] Moreover, an image input means by which this invention inputs an image for recognition in image recognition equipment which recognizes an image for recognition, A focus extract means to extract the focus from a sample image inputted with this input means, A dictionary creation means to form a partial dictionary pattern showing characteristic quantity of a partial image near [ which was extracted with this focus extract means / each ] the focus, and to draw up a partial dictionary of each focus, At the time of recognition of an image for recognition, said focus extract means extracts each focus from an input image inputted with said input means. It is image recognition equipment characterized by having a collating means to perform collating with characteristic quantity obtained from a partial image near [ each ] the focus, and a partial dictionary pattern of said partial dictionary.

[0019] Moreover, this invention is set to image recognition equipment of claim 13, and said focus extract means is image recognition equipment characterized by extracting a point with a unique brightness change in a small field in an image for recognition, or a point with a unique geometry change of brightness as the focus.

[0020] Moreover, this invention is image recognition equipment characterized by holding only a partial dictionary pattern which evaluated redundancy about two or more partial dictionary patterns obtained from a sample image, and was made into non-redundancy at the time of dictionary construction by said dictionary creation means in a partial dictionary in image recognition equipment of claim 13.

[0021] Moreover, in image-recognition equipment of claim 13, at the time of dictionary construction by said dictionary creation means, this invention do not hold two or more partial dictionary pattern \*\*\*\* obtained from a sample image, but hold a partial dictionary pattern in which characteristic quantity be show for every predetermined gap, and be image-recognition equipment characterize by complement characteristic quantity using an approximation function complement about the section in the meantime.

[0022] Moreover, it is image recognition equipment characterized by collating by this invention's creating and holding a contribution coefficient which contributes to recognition according to a use to each partial dictionary pattern in image recognition equipment of claim 13, and said collating means considering a contribution coefficient.

[0023] Moreover, this invention is image-recognition equipment characterized by to use a set of a partial dictionary pattern effective in recognition on the occasion of collating by said collating means, creating and holding a contribution coefficient which contributes to recognition according to



a use to each partial dictionary pattern, and choosing from distribution of a contribution coefficient according to a use of recognition in the image-recognition equipment of claim 13.

[0024] Moreover, this invention has a brightness frequency dictionary held in quest of brightness frequency of each pixel in an image for recognition based on dictionary data in image recognition equipment of claim 13. While collating an input image and a dictionary drawn up with said creation means at the time of recognition of an image It is image recognition equipment characterized by having having a collating means which collates by considering reliability about each pixel of an input image based on brightness of each pixel of said brightness frequency dictionary and input image in the case of this collating.

[0025] Moreover, this invention is set to an image recognition method of recognizing an inputted image for recognition. Extract the focus from an inputted sample image, form a partial dictionary pattern showing characteristic quantity of a partial image near [ this / that was extracted / each ] the focus, and a partial dictionary of each focus is drawn up. An image recognition method characterized by performing collating with characteristic quantity which extracted said focus from an inputted input image at the time of recognition of an image for recognition, and was obtained from a partial image near [ each ] the focus, and a partial dictionary pattern of said partial dictionary.

[0026] Moreover, this invention is the image recognition method characterized by setting to said focus extract and extracting a point with a unique brightness change in a small field in an image for recognition, or a point with a unique geometry change of brightness as the focus in an image recognition method of claim 20.

[0027] Moreover, this invention is the image recognition method characterized by holding only a partial dictionary pattern which evaluated redundancy about two or more partial dictionary patterns obtained from a sample image by said dictionary creation time, and was made into non-redundancy in a partial dictionary in an image recognition method of claim 20.

[0028] Moreover, in an image recognition method of claim 20, this invention does not hold two or more partial dictionary pattern \*\*\*\* obtained from a sample image to said dictionary creation time, but holds to it a partial dictionary pattern in which characteristic quantity is shown for every predetermined gap, and is the image recognition method characterized by complementing characteristic quantity using an approximation function complemented about the section in the meantime.

[0029] Moreover, it is the image recognition method characterized by collating by this invention's creating and holding a contribution coefficient which contributes to recognition according to a use to each partial dictionary pattern in an image recognition method of claim 20, and said collating means considering a contribution coefficient.

[0030] Moreover, this invention is the image-recognition method characterized by to use a set of a partial dictionary pattern effective in recognition on the occasion of collating by said collating means, creating and holding a contribution coefficient which contributes to recognition according to a use to each partial dictionary pattern, and choosing from distribution of a contribution coefficient according to a use of recognition in an image recognition method of claim 20.

[0031] Moreover, while this invention draws up a brightness frequency dictionary held in quest of brightness frequency of each pixel in an image for recognition based on dictionary data in an image recognition method of claim 20 and collating an input image and said dictionary at the time of recognition of an image It is the image recognition method characterized by having collating by considering reliability about each pixel of an input image based on brightness of each pixel of said brightness frequency dictionary and input image in the case of this collating.

[0032] Moreover, a taking-in production process which captures the target image with which this invention contains a person's face image, A focus detection production process of detecting the focus about an object image captured at the aforementioned taking-in production process based on the singular point of distribution of brightness in a predetermined field of this object image, A contribution coefficient production process of computing two or more contribution coefficient which shows effectiveness for collating about two or more set semi- data of the focus for recognition stored beforehand, respectively, and storing it, Said criteria data deletion production process of deleting two or more redundant criteria data in consideration of the redundancy of criteria data, The

remaining two or more set semi- data at the time of redundant criteria data being deleted at said criteria data deletion production process, It is the image recognition method characterized by having a collating production process which compares the focus detected at said focus detection production process for said two or more unit fields of every based on said contribution coefficient stored at said contribution coefficient production process, and collates said object image.

[0033]

[Embodiment of the Invention] Hereafter, this invention is explained to details using a drawing. A person's face field included in the image inputted from the video camera using this method is started in below, the posture is presumed to it, and two operation gestalten of the equipment for performing collating with the person in an image and the person by whom dictionary registration was done are shown in it.

[0034] (1) The outline processing explanatory drawing 1 of the 1st operation gestalt is a block diagram of the operation gestalt of this equipment. The image storage section 2 by which this person face recognition equipment is connected with the image input section 1 in drawing 1 at this, The user control unit 13 connected to this, and the dictionary creation section 14 connected to this, The brightness frequency dictionary 7 for a face field extract connected to this, the brightness frequency dictionary 8 for a face parts extract, the brightness frequency dictionary 9 for collating, the face field extract section 3, the face logging dictionary 10, the face parts extract section 4, the face parts dictionary 11, and the collating fixing section 5, It has the collating section 6, the face collating dictionary 12 connected to this collating section 6, and the display 15 connected to the collating section 6.

[0035] In order to perform actuation of above-mentioned face logging, posture presumption, and face collating, the face parts dictionaries 11, such as the \*\*\*\*\* dictionary 10 drawn up from the face image sample in two or more person object under two or more photography environment, an eye, and opening, are used for this person face recognition equipment.

[0036] Drawing 2 is the flow chart of the outline processing concerning the 1st operation gestalt concerning this invention, and explains actuation hereafter using drawing 2. The system of this operation gestalt stores in the image storage section 2 once the digital image data inputted every several frames by the image input section 1 which consists of monochrome, or the television camera and A/D converter of a color (S11). And in the face field extract section 3, the brightness reliability for a face field extract is calculated about each pixel based on the brightness frequency dictionary 7 for the face field extract sections (S13), and a face field is extracted from this brightness reliability for a face field extract, and the \*\*\*\* delivery volume dictionary 10.(S15).

[0037] Furthermore, in the face parts extract section 4, the brightness reliability for face parts inside a face field is calculated (S17), and an eye and opening are extracted from this brightness reliability and a face parts dictionary (S19). Furthermore, in the collating fixing section 5, the image for collating is generated from the location of a face parts field, and an input image (S21). And finally collating with the person in an input image and the registration person in a dictionary is performed by what (S23) pattern matching of the image for collating and a registered dictionary is computed for according to the reliability dictionary for collating in the collating section 6.

[0038] that is, although pattern matching of an input image and a face dictionary perform, it be match focusing on that whose reliability of distribution of a brightness be a constant level among the dictionary data of a face image, and the person recognition equipment which secure fixed recognition precision quick moreover realize using the brightness reliability which be the feature of the 1st operation gestalt of this invention in that case in the face field extract section 3, the face parts extract section 4, and the collating section 6.

[0039] In addition, when the input from a monochrome camera is targetted by the image input section 1 in the above, the image storage section 2 is accumulated as a gray shade image of 8bits, and when aimed at a color picture, the color component of an input image (RGB) is changed and accumulated in the brightness component Y. Generally the transformation at this time is not used and does not interfere [ following ].

[0040]

$Y=0.30, R+0.59, G+0.11, \text{ and } B$  (1) Although the above is the outline of the 1st operation gestalt, processing of the face field extract section 3, the face parts extract section 4, the collating fixing

section 5, and the collating section 6 is explained below at details.

[0041] At the beginning of creation \*\*\*\* of face logging, a collating dictionary, and a face parts dictionary (a) Two or more photography environment, Collect the sample image data of the face of two or more person object, and the field where a face is contained from these is started by the manual by the user control unit 13 which consists of a mouse etc. After setting to  $l_i$  ( $i=1, \dots, P$ , and  $P$ :all measurement size) the image vector checked by the display 15, normalizing these each in the fixed size  $N$  and normalizing the lighting fluctuation in each image, principal component analysis is carried out, the  $K$ -dimensional ( $k < N$ ) high order of characteristic value is used, and a face pattern is created. the characteristic vector ( $N$  dimension) according to the above-mentioned principal component analysis as a dictionary generated at this time — characteristic value — size — it becomes what was collected  $k$  pieces from the direction.

[0042] It is  $l_i$  so that the size of a normalization image may become fixed as normalization about size here. It is good as what was thinned out at a fixed rate. Moreover, as normalization to lighting fluctuation, "normalization by the image magnitude of a vector" (each component of a vector is broken by the norm of a vector) stated on the "conventional trouble" of this specification, "histogram flattening" [1], "the plane reliance panel of the brightness in a screen" [1], etc. may be used.

[0043] Use as the dictionary 12 for face collating what was created as sample image data for dictionary creation above only using the data of every two or more photography environment and each registrant  $j$  (however,  $j=1, \dots, P$ ;  $\sum P_j = P$ ), and let what was created using all of two or more photography environment and two or more persons' data be the face logging dictionary 10.

[0044] the face parts dictionary 11 — \*\* — although it is made like and created, only the field of an eye and opening is used, not using the whole face as a sample image.

(b) Explain a face field extract, a face parts extract, next detection of the face location in the input screen performed in the face field extract section 3 according to the flow chart shown in drawing 2. In the face field extract section 3, detection of a face location is performed about the image (subject copy) temporarily accumulated in the image storage section 2.

[0045] The luminance distribution for a face field extract is acquired from introduction and the brightness frequency dictionary 7 (S31). And the contraction image which reduced the subject copy to a certain resolution is created (S33). Next, brightness reliability for a face field extract for every screen of a contraction image is computed (S35). (updating) And the raster scan of the inside of this contraction image is carried out, a partial image is cut down according to the feature-vector size accumulated into the above-mentioned face logging dictionary 11, and matching between feature vectors is taken (S37). furthermore — and when two or more reduction percentage is changed, let a location when evaluation of the above-mentioned matching is the best be a face field (S39, S41).

[0046] As the matching technique, it does not interfere using the method of detecting so-called inner product max, or detecting so-called distance (for example, Euclidean distance, Mahalanobis \*\*\*\*\*, etc.) min about the feature vector in a dictionary, and a partial image vector so that generally as the pattern recognition technique of an alphabetic character or voice.

[0047] By the method specifically according to an inner product  $1/|V| \sum (V, D_k)$  ( $k=1, \dots, K$  however  $V$ :logging partial image-size normalization image vector,  $D_k$ : It is the  $i$ -th characteristic vector of a face logging dictionary, and  $|V|$  is the norm of  $V$ . Although  $(V, D_k)$  are the inner products of  $V$  and  $D_k$ , if both number of dimension is made into  $M$  dimensions  $(V, D_k) = \sum v(i) \times dk(i)$  (3) It is set to  $i=1$  and  $M$ . However,  $v(i)$  and  $dk(i)$  are the  $i$ -th element of each vector.

[0048] Moreover, by the method by distance, above-mentioned  $V$  and the central value  $D1$  of a dictionary sample vector, for example, the above, are projected on  $D_k$  ( $k=1, \dots, K$ ), the projection vectors  $PV$  and  $PD1$  are created, respectively, and those Euclidean distances 1 are found.

[0049] Here  $PV(k) = (V, D_k)$  (3) It is ' however  $k=1, \dots, K$ , and is  $PD1(k) = (D1, D_k)$ .

[0050] Although it is equivalent to each pixel of the image which normalized the partial image with which each element of  $V$  was cut down out of the input image in fixed size, and was created in formula (3) and formula (3) ' above, weighting is perform using the reliability judged for every pixel from the below-mentioned brightness frequency dictionary about these, and the point of this operation gestalt matches only using the pixel which will seemingly be reasonable as brightness. Therefore, (3) and (3) ' is as follows.

[0051]

$(V, D_k) = PV(k) = \sum v(i) \times T(i) \times dk(i)$

(3) " —  $i = 1$  and  $M = T(i)$  is the above-mentioned reliability here. Whenever each resolution image is generated above, brightness reliability is also updated by the below-mentioned method.

[0052] the interior of the face field detected in the above-mentioned face field extract section 3 in the input image although an eye and opening were extracted as parts in the face parts extract section 4 — receiving — this and \*\* — pattern matching [ like ] is performed using the face parts dictionary 11. Drawing 4 is a flow chart which shows extract processing of face parts. In drawing 4, one kind in a left eye, a right eye, and opening is set up as attention parts (S51). At this time, reliability is computed for every parts called an eye and opening using the separate brightness frequency for attention parts (S53). And attention parts are extracted from brightness reliability and an attention parts dictionary inside a face field (S55). Thus, attention parts are changed one after another and detection of attention parts is continued until it extracts all parts (S57) (S59).

[0053] (c) Describe creation of a brightness frequency dictionary, a reliability judging, next creation and the judgment of reliability of a brightness frequency dictionary (7, 8, 9) below.

[0054] As shown in the above-mentioned (a), the sample images about two or more registrants in two or more photography environment are collected, and the image which normalized in the fixed size N about the image  $I_i$  ( $i = 1, \dots, P$ ; P: measurement size) cut down as a face field is set to  $J_i$  ( $i = 1, \dots, P$ ; P: measurement size).

[0055] First, each  $J_i$  The-frequency distribution  $h$  of the brightness to all samples  $(x, y, I)$  is created to each pixel  $(x, y)$  of every, and it considers as a brightness frequency dictionary (7, 8, 9). And at the time of recognition, reliability  $w(x, y)$  is computed from the correspondence pixel  $I(x, y)$  of  $h(x, y, I)$  and an input partial image. As a brightness frequency dictionary, a thing separate as 7 for a face extract, 8 for a face parts extract, and 9 for face collating is prepared as follows here.

[0056] Although various things as an algorithm of the above reliability can be considered, reliability  $T(x, y)$  may be determined as follows, for example from average  $Mh$  and standard deviation  $mho$   $h$  of luminance distribution  $h(x, y, I)$ .

[0057]

$T(x, y) = \{1/\sum h(x, y, I)\} \times h(x, y, I)$

if  $|I - Mh| < \alpha \times mho$   $I = I_{min}$ , —  $I_{max} = 0$  else (4) It corrects,  $\alpha$  is the constant of the arbitration for determining the range of reliability above, and they are  $I_{max}$  and  $I_{min}$ . It is the maximum and the minimum value of a brightness value of an observation image.

[0058] Moreover, the technique by a-posteriori probability conversion as shown in reference [2] as the another technique may be used. Here, the thing and the event  $\theta_1$  which an event  $\theta_0$  is reliable shall not be reliable. When the a-priori probability of  $\theta_0$  and  $\theta_1$  before observing the brightness of the pixel in a certain partial image is  $w_0$  and  $w_1$ , respectively (however,  $w_0 + w_1 = 1$ ), a-posteriori probability  $w_1'$  — conditional [ of the pixel brightness  $I$  under event  $\theta_k$  ] — the time of setting probability to  $P(I|\theta_k)$  — from Bayes' theorem  $w_1' = \{w_1 p(I|\theta_1)\} / \{p(I|\theta_0)(1 - w_1) + p(I|\theta_1)w_1\}$  (5) It becomes. This  $w_1'$  is used as reliability.

[0059] Here, it is from the frequency distribution  $h$  of the logging image of face dictionary creation time  $(x, y, I)$ :  $p(I|\theta_0) = \{1/\sum h(x, y, I)\} \times h(x, y, I)$  (6) In order that it may be  $I = I_{min}$  and —  $I_{max}$  and  $p(I|\theta_1)$  generally may not have a data sample  $p(I|\theta_1) = 1/(I_{max} - I_{min})$  (7) It presumes. Moreover, the above updates with  $w_1 = w_1'$ , whenever an input image is updated, and it is good as initial value  $= 1.0/2.0$  of  $w_1$ .

[0060] (d) In the decision collating fixing section 5 of a collating location From the relation between the face field called for in the face field extract section 3, and the face parts called for in the face parts extract section 4 The quadrilateral field which the eye of a face and the edge of opening connect as a field with little aging as people's face is chosen, and the image for collating which deformed using the geometric conversion which this indicates is in agreement with the location of the face parts in the collating dictionary 12 to affine transformation or [3] and which starts and consists of a field is created.

[0061] (e) Use the flow chart of drawing 5 for the collating last, and explain the procedure of collating to it. Although pattern matching is performed by the technique same about the image for collating created by the collating fixing section 5 as the collating dictionary 12 and (b) having shown,

in this case, it is every registrant's  $j$  dictionary  $D_j$  in the collating dictionary 12 as a candidate for collating. It becomes.

[0062] That is, Individual's  $i$  thing  $D_i$  is set up as an attention registered person's dictionary (S61). And an attention person's brightness reliability is calculated (S63). Furthermore, similarity (distance) with an attention person is computed from illuminance reliability and  $D_i$  inside a collating field (S65). Also in here, matching based on (3) " types using the reliability judged about each pixel within a logging image is performed. Here, they are an image for collating, and Dictionary  $D_j$ . In matching, since it differs for every individual dictionary, brightness frequency distribution may use different distribution for every dictionary for collating, may create one frequency distribution of the whole registrant, and may use only this.

[0063] Furthermore, collating is completed with outputting ID of the person to whom this was carried out until it collated this matching with all registrants (S67, S71), and optimal matching was carried out (S69).

[0064] Furthermore, although it is how to treat the reliability in the case of this matching, that at least two method is considered. According to the degree of reliability, one makes this value a coefficient and it matches according to this magnitude. That is, about the unit field (mesh) where reliability is high, matching near 100% is performed with a subject-copy image in dictionary data. On the other hand, the degree of matching with a low degree is reduced about the unit field (mesh) where reliability is low. By carrying out like this, the optimal matching according to the degree of brightness is realizable.

[0065] Another prepares the threshold about reliability, and if this reliability is a value below constant value, it will not perform matching with a subject-copy image about the unit field (mesh) of the dictionary data about this reliability. With constant value [ more than ], the unit field (mesh) of this dictionary data performs matching with a subject-copy image.

[0066] (2) Use a drawing for below and explain the 2nd operation gestalt of outline processing explanation of the 2nd operation gestalt, next this invention to it at details.

[0067] The feature of the 1st operation gestalt searches for the reliability about the brightness of dictionary data, it is treating dictionary data according to this reliability, and without dropping an essential discernment precision, is reducing the frequency of matching and raises the speed of recognition processing.

[0068] The so-called contribution coefficient which, on the other hand, expresses the degree which contributes to discernment processing among each unit field (mesh) of dictionary data as the 2nd point of this invention in the 2nd operation gestalt is calculated, and this is made to correspond to each unit field of each dictionary data, and is made to memorize beforehand. And about the unit field where a contribution coefficient is low, although it corresponds to this other than the face field on a screen (for example, the image of a background), in the case of matching with a subject-copy image, as this does not consider as the object of matching, it shortens the time amount concerning matching, and attains speeding up of recognition processing, for example.

[0069] Drawing 6 is the block diagram of the 2nd operation gestalt. The image storage section 2 by which this person face recognition equipment is connected with the image input section 1 in drawing 6 at this, The focus extract section 18 connected to this, and the dictionary creation section 19 connected to this, The dictionary reevaluation section 20 connected to this, and the partial feature dictionary 23 connected to this, The face field extract section 21 connected to this, and the face posture presumption section 22 connected to this, The collating section 6 connected to this, and the contribution coefficient 24 for a face field extract connected to the dictionary creation section 19, the contribution coefficient 25 for face posture presumption and the contribution coefficient 26 for collating, Furthermore, it has the brightness frequency dictionary 27 for a face field extract connected to the dictionary creation section 19, the brightness frequency dictionary 28 for face posture presumption, the brightness frequency dictionary 29 for collating, the display 31 connected to the collating section 60, and the user control unit 30 connected to this, respectively.

[0070] In such structure, the contribution coefficient the person face recognition equipment of the 2nd operation gestalt indicates the recognition effectiveness of each projection data in a dictionary to be for every use of a face field extract, face posture presumption, and collating is created by the dictionary creation section 19. Furthermore, speeding up of collating processing is attained, the unit

field (mesh) of dictionary data not being used for matching for discernment, or not being used, and holding a substantial discernment precision according to the degree of this contribution coefficient. [0071] Hereafter, actuation of processing of the 2nd operation gestalt and the point of other this inventions are explained using the flow chart of drawing 7 and drawing 8. The system of this operation gestalt stores in the image storage section 17 once the digital image data inputted every several frames by the image input section 16 which consists of monochrome, or the television camera and A/D converter of a color (S71), cuts down the image of the perimeter of the focus obtained with the focus extract means 18, and draws up the above-mentioned dictionaries 23, 27, 28, and 29 with the dictionary creation means 19.

[0072] First, the digital image inputted by the image input section 16 every several frames is inputted, and it stores in the image storage section 17. Furthermore, as the 3rd point of this invention, in order to perform actuation of face logging, posture presumption, and face collating, the focus extract section 18 extracts the focus effective in recognition from the face image sample in two or more person object under two or more photography environment efficiently (S73). And in the dictionary creation section 19, it has a small field near the focus as an aperture, and the subspace which carried out principal component analysis of these apertures image, and the projection vector which projected each aperture to the space are created as a dictionary 24 (S75). according to this method — an input image — also receiving — \*\* — a part of face field according to partial lighting fluctuation and other bodies of the face image at the time of recognition by calculating the projection to said subspace of the image [ like ] in window region, and using matching with them and the projection in a dictionary — it also receives hiding and generating of incorrect recognition can be controlled.

[0073] Furthermore, as the 4th point of this invention, evaluation and compression of the redundancy of the projection data in a dictionary are performed in the dictionary reevaluation section 20. That is, when projection data draws the same pattern as a repeat and exists in dictionary data, in case matching with a subject-copy image is performed, by deleting data which does not not much have comparative semantics, and data with high redundancy, dictionary data is reduced and, thereby, matching speed is raised (S77).

[0074] Next, the approximation function expression of the image projection fluctuation inside each sample raised as the 5th point of this invention is performed, and this performs efficient reduction of dictionary data (S79). That is, when dictionary data can express with an approximation function, the amount of data can be sharply reduced by expressing this with an approximation function. Thereby, matching speed can be raised very much.

[0075] Such an image sample is classified according to a manual for every use (S81). And in the dictionary creation section 19, the contribution coefficient which is the feature as the 2nd point of this invention raised first is computed. That is, the contribution coefficient showing the degree which participates in discernment processing among each unit field (mesh) of dictionary data is calculated, and this is made to correspond to each unit field of each dictionary data, and is made to memorize beforehand (S83). And about the unit field where a contribution coefficient is low, although it corresponds to this other than the face field on a screen (for example, the image of a background), in the case of matching with a subject-copy image, as this does not consider as the object of matching, it shortens the time amount concerning matching, and attains speeding up of recognition processing, for example.

[0076] And improvement in processing speed is aimed at by choosing from the statistical distribution over all dictionary samples only what has a high contribution coefficient as the 6th point of this invention more than a threshold, and using it for matching (S85).

[0077] Next, from an input image, in the face field presumption section 21, the face posture presumption section 22, and the collating section 23, a face exists where in a screen and it asks for whose face it is with what kind of sense from the result of matching with the partial image of the perimeter of the focus extracted in the focus extract section 18, and the projection data in a dictionary. Matching which carried out with [ which followed the contribution coefficient in the above-mentioned dictionary in these the processings of each ] weight is used.

[0078] moreover, the above matching — setting — further — as the 7th point of this invention — the 1st operation gestalt and \*\* — technique [ like ] is applied, brightness frequency is computed

by the dictionary creation section 19, and it memorizes in the brightness frequency dictionary (27, 28, 29). And by computing and using the reliability based on brightness frequency distribution, matching by which dictionary data was refined very much is realized, and only recognition speed will improve, with recognition precision held.

[0079] in addition, the 1st operation gestalt and \*\* — like, when the input from a monochrome camera is targetted by the image input section 1, the image storage section 2 is accumulated as a gray shade image of 8bits, and when aimed at a color picture, the color component of an input image (RGB) is changed and accumulated in the brightness component Y according to (1) type.

[0080] Next, drawing 8 is a flow chart explaining the actuation at the time of recognizing a face image, memorizes first the face image which is a very [ recognition ] image in the image storage section 17 by the input section 16 (S91), then, extracts the focus in an input face image in the focus extract section 18, and cuts down the partial image of the perimeter (S93). Furthermore, matching between the partial images and dictionary data which were cut down in the face field extract section is performed, and a face field is presumed (S95). In this case, the contribution coefficient 24 for a face field extract and the brightness frequency based on the brightness frequency dictionary 27 for a face field extract are used together, and a face field is presumed. Furthermore, matching with the partial image and dictionary 23 based on the coefficient 25 for face posture presumption and the brightness reliability computed based on the brightness frequency dictionary 28 for face posture presumption is performed, and a face posture is presumed (S97).

[0081] Furthermore in the collating section 60, matching with a partial image and a dictionary 23 is performed based on the contribution coefficient 26 for collating, and the brightness reliability computed based on the brightness frequency dictionary for collating, and it is judged who the face image was (S99).

[0082] Next, the details of "an extract of the focus", "construction of a dictionary and compression", and the "matching method" are explained among the processings mentioned above. In the above, the outline of processing actuation of the 2nd operation gestalt was explained.

[0083] (a) By extracting the feature portion of an image which are the focus extract from an input image and also the 3rd point of this invention explains below the production process which performs image recognition at details. The focus extract section 18 extracts the feature effective in recognition from the inside of an image by the following methods as the 3rd point of this invention.

[0084] In case matching of an input image and a dictionary image, continuation inter-frame matching, etc. are performed, it is common to choose a certain field and to make the perimeter into a template. Although it is for one improvement in the speed of this reason, another is for choosing a point, i.e., other points and the point that distinction is attached, more effective in matching (based mainly on distance or correlation) compared with the case where all the pixels equivalent to an object are made into a trace processing object. The edge in an image is well known as such a thing.

[0085] such [ here ] a texture — in order to search for the unique effective feature, the effective focus extract technique of [4] is used, and it accelerates with pretreating to this. About the image for dictionary creation, or the whole input image surface for recognition, when processing [4], it is necessary to calculate the characteristic value of 2x2 matrix about each pixel, and processing cost starts. Then, beforehand, all the textures in a processing field extract a corner point by the below-mentioned method as a characteristic point geometrically, and limit an effective focus extract field only near the corner point.

[0086] The production process of the following processings is explained based on drawing 9. The corner appearance in here evaluates the edge it is supposed that it is good characteristic quantity experientially in the mask size of 5x5 pixels, and judges a corner point.

[0087] First, if an image is inputted by the image input section 16 (S101), sobel [5] will be beforehand applied to a subject copy, and discriminant analysis binary-ization [6] will be performed (S103). And about each pixel in this image, when an attention pixel is 1, convolution with the mask of drawing 10 (a) is taken. And with the value of the pixel value p of a result, it judges whether an observing point is close to "+" mold branch point, or it is close to "x" mold branch point. here — as follows — a case — having divided .

[0088]

the  $p \geq 2 \times \text{maximum pixel value} \rightarrow \text{"+" mold branch point}$  The  $p \leq -2 \times \text{maximum pixel value} \rightarrow \text{"x"}$



mold branch point Except the above  $\rightarrow$  the degree which is not a corner candidate — “ — the case where it is based on the mask of drawing 10 (b) about “+” mold and each “x” mold — dividing — coming out — the above and \*\* — it is made like and a corner point is searched for (S namely, the time of a “+” mold — as follows — a part for a case — an opium poppy — if it is not a  $\rightarrow$  corner point — a case — dividing — carrying out — . The — maximum pixel value The  $\leq p <$  maximum pixel value  $\rightarrow$  Corner point Except the above  $\rightarrow$  Case of “x” mold branch point which is not a corner point The — maximum pixel value The  $\leq p <$  maximum pixel value  $\rightarrow$  Corner point Except [0089] According to a degree type, the focus is judged about the inside of the perimeter field W of the observing point p (x y) finally judged to be a corner (S107). It is Gradient of the following first. matrix It asks for G.

[0090]

$$G = \text{sigma}(gx * gx \quad gy * gy) = (a \ b) \begin{pmatrix} 8 \\ W \end{pmatrix} (gy * gx \quad gy * gy) \begin{pmatrix} c \\ d \end{pmatrix}$$

Next, it asks for characteristic value  $\lambda^+$  of G, and  $\lambda^-$ , and let the larger one of them be the pixel value of an observing point. A certain threshold  $\lambda_{\text{dat}}$  after following all the pixels of a processing field in the above-mentioned processing  $\lambda > \lambda_{\text{dat}}$  (9) A part for the becoming pixel value part is extracted as the effective focus. That for which it asked by discriminant analysis about each pixel value inside a processing field is used for the threshold at this time.

[0091] (b) In construction of a dictionary and the compression dictionary creation section 19 About each face image sample, the partial image inside the field of the perimeter WxH pixel of the recognition effective focus extracted by the method of of the foregoing paragraph (2) and (a) is cut down. Principal component analysis of them is carried out, and the projection of a fractionation image vector is all searched for about the matrix (dictionary space matrix) which asked even for the Kth characteristic vector of them, and all the face samples to the space, and it considers as a dictionary.

[0092] However, since the above-mentioned dictionary becomes huge when there are many measurement sizes, dictionary capacity is reduced according to the 4th point of above-mentioned this invention, and the 5th point. First, with the method as the 4th point of this invention, it is the projection  $P_i$  to the dictionary space of each partial image  $W_i$  ( $i = 1, \dots, N_i$ ) inside certain sample image I. It asks and they are its projection among the partial images in the other sample image J ( $J = 1, \dots, N_j$ ), and  $P_j$ . Distance is min so-called  $[w_j]$ . The distance  $P_j$  It asks. By this, it is  $w_j$ . Since the sample image which belongs is called for, it is the partial image  $W_i$  in Image I. It matches and also the frequency distribution of a sample image is acquired. this time — frequency distribution — max — other sample image  $J_{\text{max}}$  It is defined as having matched Image I. Moreover, I and  $J_{\text{max}}$  at this time Distance  $D_{ij}$  is defined below.

[0093]

$D_{ij} = 1/j \sum |P_i - P_j| \quad (10) \quad j = 1, \dots, K$ , however  $P_j$  : A thing, K from which a partial image and projection distance in Image I serve as min among the partial images in Image J: It is the number. Moreover, Euclidean distance and absolute value distance may be used as a distance.

[0094] The example of the table of the result of having calculated the above about all samples is shown in drawing 11 . The frequency distribution of such distance ( $D_{12} - D_{nn-1}$  in drawing) is created, and with the application of a discriminant analysis method [6], a threshold is calculated so that the variance ratio of frequency distribution may be made into max. At this time, that projection is excepted from a dictionary about the partial image in the sample image whose distance was nearer than the threshold.

[0095] Moreover, by the method as the 5th point of this invention, it shall apply, after performing dictionary reduction per sample image by the method of the 3rd term. First, the face image sample for dictionary creation is sorted about a certain individual in order of a certain parameter, for example, lighting, and fluctuation of a direction. If it does in this way, it will become general and the projection of the partial image in between [ which suits ] sample images will continue.

[0096] a part for next, all of above-mentioned sample images — an image — \*\*\*\*\* — although the projection distance between each partial image is found — if — a certain partial image  $w_i$  from — partial image  $w_j$  of others when it sees the case where the minimum distance is suited — coincidence —  $w_j$  from — a minimum distance partial image —  $w_i$  it is — sometimes, it is considered that both are resemblance. And when the partial images  $w_{i1}$ ,  $w_{i2}$ , and  $w_{i3}$  in adjoining



sample image sequences and —win are similar, an approximation function expresses these collectively and dictionary size is reduced. Although a \*\*\*\* thing [ like ] can be considered as an approximation function, the cubic spline used for [7] etc. may be used.

[0097] Furthermore, as the 2nd point of this invention, according to the use of face extract posture presumption and collating, the contribution coefficient for effective recognition is created from a dictionary sample, and adjustment of the method of effect in the recognition result of a partial dictionary pattern is enabled for every use about each partial dictionary pattern in the dictionary creation section 19 at the time of recognition. The example of the calculation method of a contribution coefficient is raised to below.

[0098] It asks for the above-mentioned focus image about two or more samples first arranged for every use. In order that the luminance distribution of this image may show singularity within a sample image, for the optimal recognition, there should be recognition contribution for that use in that partial—near the focus image highly. Then, principal component analysis of said focus image is carried out for every use sample, and the characteristic value  $P_i$  of  $K$  high orders and a characteristic vector  $V_i$  ( $i=1 \sim n$ ) are extracted.  $\sum P_i \times V_i$  (11) The weight image for every use is created by  $i=1$  and  $k$ .

[0099] Each partial image  $w_i$  at the time of putting each sample image  $I$  on the above-mentioned weight image finally, and making it meet The contribution coefficient  $C_{wi}$  for every partial image is computed as follows by the total  $S$  of the weight pixel value in a field.

[0100]

$$C_{wi}=S/P_{max} \quad (12)$$

It is  $P_{max}$  here. It is the maximum pixel value in the above-mentioned weight image. In addition, it is made to be the following as a sample for every use, for example.

[0101] In the case of face field presumption : Thing face posture presumption which set the value outside a face field to 0 by the feature image of each sample image : Face collating same as the above : what created the feature image about the sample image set for every individual — an individual contribution coefficient by creation and the above for every individual dictionary in this case Although effect of the recognition processing on a partial image dictionary with high redundancy is lessened if possible and the purpose improves precision with the contribution coefficient for every use according to a use In order to prevent the processing-time increase by matching in a partial image dictionary with a still lower contribution, as already stated as the 6th point of this invention, a threshold is statistically determined from distribution of these coefficients, and if the partial image dictionary of the contribution coefficient below the threshold is not used, it is good. for this reason — contribution coefficient  $C_{wi}$  which took the frequency distribution of the contribution coefficient called for for every above-mentioned partial image dictionary, calculated the threshold  $CT$  by the discriminant analysis method [6], and was given for every partial image dictionary  $C_{wi} < CT$  (13) — a partial image dictionary is not used for matching.

[0102] (c) Match in the following procedures about the inside of an input image in the matching method face field presumption section 21, the face posture presumption section 22, and the collating section 23.

[0103] the partial image dictionary which extracted the partial image from the inside of a Production process 1:input image, and was registered into the dictionary about those each — it asks for so-called similarity min inside (matching).

Production process 2: Ask for the thing  $I$  which has the highest frequency of occurrence, and the frequency  $H$  at that time among the images for samples with which the above-mentioned matching partial dictionary pattern is contained.

[0104] Production process 3: Ask for reliability  $T=$  (the total number of partial windows in maximum frequency-of-occurrence  $H$  / input image).

the partial window equivalent to the Production process 4:above  $I$  — the shift of a dictionary to a coordinate — respectively — asking — the frequency of the vote operation to a shift amount — max — a window is chosen and it asks for the affine transformation from them and a dictionary.

[0105] In the above, the collating section 23 performs processing of a production process 1, a production process 2, and a production process 3, and the face posture presumption section 22 performs a production process 1 — a production process 4. Finally, following a production process 1

— a production process 4, to the face image field in the sample image I, above-mentioned affine transformation is performed and the face field within an input image is presumed in the face field presumption section 21.

[0106] In addition, I in the above is made into estimate and it rejects with assessed reliability T and an affine transformation coefficient. the contribution coefficient C for every use stated in all knots in the production process 1 above — using — (3) " and \*\* — like — partial image I<sub>i</sub> Projection P<sub>i</sub> to dictionary space If it asks for (k), matching in consideration of a contribution coefficient can be performed.

[0107]

$P_i(k) = \sum_j I(j) \times C(i) \times dk(j)$  (14) For  $j = 1, M$ , however M, the number of dimension of a partial image vector and C(i) are the partial image I<sub>i</sub>. A contribution coefficient and dk(j) are the No. j elements of the k-th characteristic vector of dictionary space.

[0108] It asks like. further — the reliability of the brightness as the 7th point of this invention — (1), (c), and \*\* — the time of recognition — an input image to the above, and \*\*, if a partial image is cut down like, it asks like T(x y) of (4) types, or w1' of (5) types from each coordinate in the field (x y) and it is set to T(i)  $P_i(k) = \sum_j I(j) \times C(i) \times dk(j) \times T(i)$  (15) The robust recognition which took the reliability of brightness into consideration by carrying out like  $j = 1$  and M in addition to the recognition contribution for every use is attained.

[0109] As mentioned above, according to the person recognition equipment of this invention, the following effects can be considered. that is, although the image recognition \*\*\*\* method be propose as what do not use not much special equipment in that case although a user have the needs of the individual collating section with few burdens by non-contact, with equipment, there be a defect that exact recognition be difficult, according to an image recognition conventionally in the environment which do not have little lighting fluctuation in the field which went as the password of OA equipment, such as individual collating with social systems, such as close recession management, an ATM device, and a ticket machine, and a personal computer. On the other hand, according to this invention, the reliability of the luminance distribution in a dictionary creation sample image is verified about each pixel brightness in an input image, and by using only a pixel reliable about brightness, or the pixel of the circumference of it for recognition, even if there is a big change of brightness, it makes it possible to perform stable recognition.

[0110] moreover, it does not consider as the image of the whole field which has a face as a face dictionary, but creates using the subregion near [ which is included in a face image ] the partial feature field (partial dictionary pattern), and those matching results are depended on some brightness fluctuation and other bodies by carrying out a comprehensive judgment — hiding — etc. — the fluctuation to depend is absorbable.

[0111] Under the present circumstances, it poses a problem that the data size of said partial dictionary pattern becomes large, and the equipment scale for realizing this technique becomes large. However, the redundancy for every dictionary creation sample image is evaluated statistically, a redundant pattern is removed or dictionary size can be made small by approximating and holding each partial dictionary pattern with a complement [ not itself but ] function between samples which characteristic quantity change follows etc.

[0112] Furthermore according to each processing called a face extract, posture presumption, and collating, it becomes possible by using it in consideration of contribution of the optimal partial pattern to improve the face recognition engine performance remarkably from both sides of the precision of recognition, and the processing time.

[0113] In addition, bibliography is shown below.

[1] KSung and T.Poggio. and Example-based Learning for View-based Human FaceDetection.In Proceedings from Image Understanding Workshop, pp.843-850, and November 1994. [2] Nakai: The migration body detection technique and \*\*\*\*\* using a-posteriori probability, 94-valve flow coefficient-90, pp.1-8, and 1994[3] Yamaguchi, Fukui, Maeda, The face recognition system, Shingaku Giho using a dynamic image, PRMU 97-50, pp 17-23, [(1997) 4] Shi, Tomasi, Good Features to Track, and IEEE Conference on Computer Visionand Pattern Recognition(CVPR94) Seattle and June 1994 [5] Tree, Shimoda, a "image-analysis handbook", University of Tokyo Press, and pp. — 553-554 and 1991[6 —] Otsu — Distinction and the automatic threshold selecting method based on the

minimum square criteria, \*\*\*\*\*, Vol.J63-D, No.4, pp.349-356, and 1980[6] Tree, Shimoda: Image-analysis handbook PP.578[7] Murase, NEIYA, image spotting of the three-dimension body by "multiplex resolution and proper space expression, \*\*\*\*\*, Vol.36 No.10, pp 2234-2243, Oct, 1995 [0114]

[Effect of the Invention] By as mentioned above, the thing for which according to this invention the reliability about brightness is searched for from the dictionary data for performing recognition processing of a face image, and this is associated and stored as explained to details Even if it uses or uses for matching processing with the subject-copy image from a camera to the unit field (mesh) of the dictionary data which seldom fits recognition when the lighting at the time of for example, dictionary data being created on the occasion of recognition processing is unsuitable, matching processing with a low degree is performed. Matching is ensured about the mesh which realizes positive recognition by this, and the mesh with low reliability can realize discernment processing of a quick person's face image as a whole by carrying out an abbreviation etc., holding discernment precision.

[0115] moreover — further — the unit field (mesh) of dictionary data — it is alike, respectively, and attaches, and a contribution coefficient is prepared from a viewpoint whether to contribute to matching processing however, respectively, and effective [ according to this contribution coefficient ] in the case of matching — two or more meshes are alternatively matched as criteria. Thereby, discernment processing of a quick person's face image is realizable by reducing dictionary data, holding discernment precision.

---

[Translation done.]

**\* NOTICES \***

**Japan Patent Office is not responsible for any damages caused by the use of this translation.**

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

---

**DESCRIPTION OF DRAWINGS**

---

**[Brief Description of the Drawings]**

**[Drawing 1]** The whole block diagram concerning the 1st operation gestalt of this method.

**[Drawing 2]** The flow chart of the outline processing concerning the 1st operation gestalt of this invention.

**[Drawing 3]** The flow chart of the face field extract concerning the 1st operation gestalt of this invention.

**[Drawing 4]** The flow chart of an extract of the face parts concerning the 1st operation gestalt of this invention.

**[Drawing 5]** The flow chart of the face collating processing concerning the 1st operation gestalt of this invention.

**[Drawing 6]** The whole block diagram concerning the 2nd operation gestalt of this method.

**[Drawing 7]** The flow chart of dictionary creation with the 2nd operation gestalt of this invention.

**[Drawing 8]** The flow chart of recognition with the 2nd operation gestalt of this invention.

**[Drawing 9]** The focus extract technique concerning the 2nd operation gestalt of this invention.

**[Drawing 10]** The corner extract mask concerning the 2nd operation gestalt of this invention.

**[Drawing 11]** The example of distance between the sample images concerning the 2nd operation gestalt of this invention.

**[Description of Notations]**

- 1 — Image input section
- 2 — Image storage section
- 3 — Face field extract section
- 4 — Face parts extract section
- 5 — Collating fixing section
- 6 — Collating section
- 7 — Brightness frequency dictionary for a face field extract
- 8 — Brightness frequency dictionary for a face parts extract
- 9 — Brightness frequency dictionary for collating
- 10 — Face logging dictionary
- 11 — Face parts dictionary
- 12 — Face collating dictionary
- 13 — User control unit
- 14 — Dictionary

---

**[Translation done.]**

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2000-99722

(P2000-99722A)

(43)公開日 平成12年4月7日(2000.4.7)

(51) Int.Cl.<sup>7</sup>  
G 0 6 T 7/00

識別記号

F I  
G O 6 F 15/62

テーマコード・(参考)

4 6 5 K      5 B 0 4 3

審査請求 未請求 請求項の数27 OL (全 16 頁)

(21)出願番号 特願平10-268419

(22) 出願日 平成10年9月22日(1998.9.22)

(71)出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72) 発明者 榎本 暢芳

神奈川県川崎市幸区柳町70番地 株式会社  
東芝柳町工場内

(74) 代理人 100058479

弁理士 鈴江 武彦 (外6名)

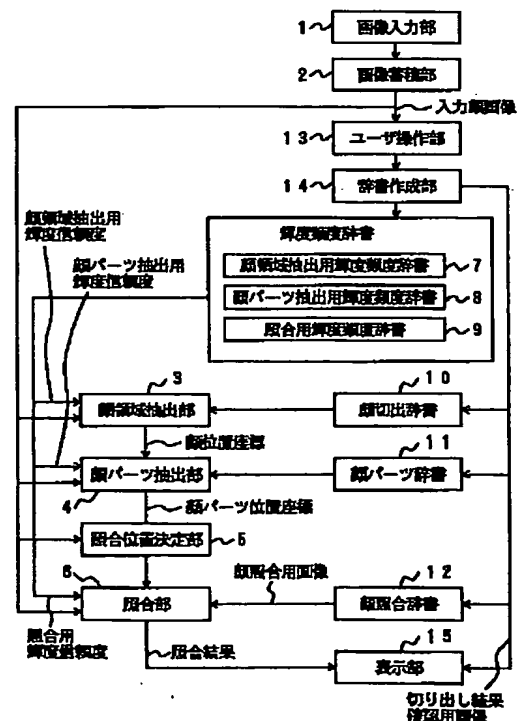
Fターム(参考) 5B043 AA09 BA04 GA05 GA17

(54)【発明の名称】 人物顔認識装置及び人物顔認識方法

(57) 【要約】

【課題】 人物の顔の画像に関して迅速で正確な認識処理を行う人物顔認識装置とこの人物顔認識方法を提供する。

【解決手段】 対象の画像を取り込む取込部１と、予め保存された輝度頻度に関する辞書データ７，８，９とこの対象画像の複数の単位領域ごとの輝度とを比較して、対象画像の複数の単位領域ごとの信頼度を求める信頼度計算部７，８，９と、予め保存された認識のための顔照合辞書１２と対象画像とを、算出された信頼度に基づいて単位領域ごとに比較し、対象画像を照合する照合部６とを有する画像認識装置。



(2)

## 【特許請求の範囲】

【請求項1】認識対象画像の認識を行なう画像認識装置において、

認識対象画像を入力する画像入力手段と、

この入力手段にて入力された認識対象画像から照合用の辞書を作成する辞書作成手段と、

この作成手段にて作成した複数の辞書データに基づき認識対象画像における各画素の輝度頻度を求めて保持する輝度頻度辞書と、

認識対象画像の認識時には入力画像と前記作成手段にて作成された辞書とを照合するとともに、この照合の際には前記輝度頻度辞書と入力画像の各画素の輝度に基づき入力画像の各画素について信頼度を加味して照合を行なう照合手段とを有することを有することを特徴とする画像認識装置。

【請求項2】請求項1に記載の画像認識装置において、前記照合手段は前記輝度頻度辞書に基づき認識対象画像として尤もらしくない画素については照合に用いないようにしたことを特徴とする画像認識装置。

【請求項3】請求項1に記載の画像認識装置において、前記照合手段は前記輝度頻度辞書に基づき認識対象画像として尤もらしくない画素については照合結果に対する寄与度を低くするようにしたことを特徴とする画像認識装置。

【請求項4】請求項1に記載の画像認識装置において、前記輝度頻度辞書は前記作成手段による辞書データ作成時の切出し画像の輝度頻度を求めて保持する顔領域抽出用輝度頻度辞書を有し、前記抽出手段は前記輝度頻度辞書に基づき認識対象画像の切出しを行なう顔領域抽出手段を有することを特徴とする画像認識装置。

【請求項5】請求項4に記載の画像認識装置において、前記辞書作成手段は、認識対象画像中の特定領域のパーツ画像に関するパーツ辞書を作成し、前記前記輝度頻度辞書は前記作成手段によるパーツ辞書作成時のパーツ画像の輝度頻度を求めて保持する類パーツ抽出用輝度頻度辞書を有し、前記抽出手段は前記輝度頻度辞書に基づき前記顔領域抽出手段にて抽出された認識対象画像から類パーツ画像の切出しを行なう類パーツ抽出手段を有することを特徴とする画像認識装置。

【請求項6】請求項5に記載の画像認識装置において、前記辞書作成手段は認識対象画像中の特定領域の認識対象画像に関する顔照合辞書を作成し、前記前記輝度頻度辞書は前記作成手段による顔照合辞書の認識対象画像の輝度頻度を求めて保持する顔照合用輝度頻度辞書を有し、前記抽出手段は類パーツ抽出手段にて抽出された類パーツ画像との関係から入力画像中の特定領域の認識対象画像を照合用画像として選択する照合位置決定手段を有し、前記照合手段は顔照合辞書と照合用画像とを照合する際に前記顔照合用輝度頻度辞書に基づき基づき認識対象画像として尤もらしい画素を用いて照合を行なうよ

2

うにしたことを特徴とする画像認識装置。

【請求項7】認識対象画像の認識を行なう画像認識方法において、入力された認識対象画像から照合用の辞書を作成し、この作成した複数の辞書データに基づき認識対象画像における各画素の輝度頻度を求めて保持しておき、認識対象画像の認識時には入力画像と前記辞書とを照合するとともに、この照合の際には前記輝度頻度辞書と入力画像の各画素の輝度に基づき入力画像の各画素について信頼度を加味して照合を行なうようにしたことを特徴とする画像認識方法。

【請求項8】請求項7に記載の画像認識方法において、前記入力画像と前記辞書とを照合する際には前記輝度頻度辞書に基づき認識対象画像として尤もらしくない画素については照合に用いないようにしたことを特徴とする画像認識装置。

【請求項9】請求項7に記載の画像認識方法において、前記入力画像と前記辞書とを照合する際には前記輝度頻度辞書に基づき認識対象画像として尤もらしくない画素については照合結果に対する寄与度を低くするようにしたことを特徴とする画像認識装置。

【請求項10】認識対象画像の認識を行なう画像認識方法において、入力された画像から認識対象画像領域を切出し、切出された認識対象画像領域の画像中の特定領域の認識対象画像から照合用の顔照合辞書を作成し、かつ切出された認識対象画像領域における各画素の輝度頻度を求めて顔領域抽出用輝度頻度辞書を作成し、認識対象画像の認識時には入力画像中の画像のうち前記輝度頻度に基づき認識対象画像として尤もらしい領域を認識対象画像として抽出し、この抽出された認識対象画像中の特定領域の認識対象画像から照合用画像を選択して前記顔照合辞書と照合するようにしたことを特徴とする画像認識方法。

【請求項11】請求項10に記載の画像認識装置において、前記顔照合辞書を作成する際に、認識対象画像中の特定領域のパーツ画像に関するパーツ辞書を作成するとともに、パーツ辞書作成時のパーツ画像の輝度頻度を求めてパーツ抽出用輝度頻度辞書を作成し、前記顔照合辞書と入力画像との照合を行なう場合には、前記顔領域抽出用輝度頻度辞書に基づき抽出された認識対象画像からパーツ抽出用輝度頻度辞書及びパーツ辞書に基づき入力画像から特定領域のパーツ画像を抽出し、抽出されたパーツ画像の位置関係から照合用画像を入力画像から選択して前記顔照合辞書と照合するようにしたことを特徴とする画像認識方法。

【請求項12】請求項10に記載の画像認識方法において、前記顔照合辞書を作成する際に、顔照合辞書の認識対象画像の輝度頻度を求めて保持する顔照合用輝度頻度辞書を作成有し、前記顔照合辞書と照合用画像とを照合する際には前記顔照合用輝度頻度辞書に基づき認識対象画像として尤もらしい画素を用いて照合を行なうように

(3)

したことを特徴とする画像認識方法。

【請求項13】認識対象画像の認識を行なう画像認識装置において、

認識対象画像を入力する画像入力手段と、

この入力手段にて入力されたサンプル画像から特徴点を抽出する特徴点抽出手段と、

この特徴点抽出手段にて抽出された各特徴点近傍の部分画像の特徴量を表す部分辞書パターンを形成して各特徴点の部分辞書を作成する辞書作成手段と、

認識対象画像の認識時には前記入力手段にて入力された入力画像から前記特徴点抽出手段にて各特徴点を抽出して各特徴点近傍の部分画像から得られた特徴量と前記部分辞書の部分辞書パターンとの照合を行なう照合手段とを有することを特徴とする画像認識装置。

【請求項14】請求項13の画像認識装置において、前記特徴点抽出手段は認識対象画像中の小領域内における輝度変化が特異な点又は輝度の幾何学的変化が特異な点を特徴点として抽出することを特徴とする画像認識装置。

【請求項15】請求項13の画像認識装置において、前記辞書作成手段による辞書構築時に、サンプル画像から得られた複数の部分辞書パターンについての冗長性を評価し、非冗長とされた部分辞書パターンのみを部分辞書に保持することを特徴とする画像認識装置。

【請求項16】請求項13の画像認識装置において、前記辞書作成手段による辞書構築時に、サンプル画像から得られた複数の部分辞書パターン金てを保持するのではなく、所定の間隔ごとに特徴量を表す部分辞書パターンを保持しておき、その間の区間については補完する近似関数を使用し特徴量を補完することを特徴とする画像認識装置。

【請求項17】請求項13の画像認識装置において、各部分辞書パターンに対して用途に応じて認識に寄与する寄与係数を作成して保持しておき前記照合手段は寄与係数を加味して照合を行なうようにしたことを特徴とする画像認識装置。

【請求項18】請求項13の画像認識装置において、各部分辞書パターンに対して用途に応じて認識に寄与する寄与係数を作成して保持しておき、前記照合手段による照合の際には、認識の用途に応じて寄与係数の分布から認識に有効な部分辞書パターンの集合を選択して使用するようにしたことを特徴とする画像認識装置。

【請求項19】請求項13の画像認識装置において、辞書データに基づき認識対象画像における各画素の輝度頻度を求めて保持する輝度頻度辞書を有し、画像の認識時には入力画像と前記作成手段にて作成された辞書とを照合するとともに、この照合の際には前記輝度頻度辞書と入力画像の各画素の輝度に基づき入力画像の各画素について信頼度を加味して照合を行なう照合手段とを有することを特徴とする画像認識装置。

4

【請求項20】入力された認識対象画像の認識を行なう画像認識方法において、入力されたサンプル画像から特徴点を抽出し、この抽出された各特徴点近傍の部分画像の特徴量を表す部分辞書パターンを形成して各特徴点の部分辞書を作成しておき、認識対象画像の認識時には入力された入力画像から前記特徴点を抽出して各特徴点近傍の部分画像から得られた特徴量と前記部分辞書の部分辞書パターンとの照合を行なうようにしたことを特徴とする画像認識方法。

10 【請求項21】請求項20の画像認識方法において、前記特徴点抽出においては認識対象画像中の小領域内における輝度変化が特異な点又は輝度の幾何学的変化が特異な点を特徴点として抽出することを特徴とする画像認識方法。

【請求項22】請求項20の画像認識方法において、前記辞書作成時に、サンプル画像から得られた複数の部分辞書パターンについての冗長性を評価し、非冗長とされた部分辞書パターンのみを部分辞書に保持することを特徴とする画像認識方法。

20 【請求項23】請求項20の画像認識方法において、前記辞書作成時に、サンプル画像から得られた複数の部分辞書パターン金てを保持するのではなく、所定の間隔ごとに特徴量を表す部分辞書パターンを保持しておき、その間の区間については補完する近似関数を使用し特徴量を補完することを特徴とする画像認識方法。

30 【請求項24】請求項20の画像認識方法において、各部分辞書パターンに対して用途に応じて認識に寄与する寄与係数を作成して保持しておき前記照合手段は寄与係数を加味して照合を行なうようにしたことを特徴とする画像認識方法。

【請求項25】請求項20の画像認識方法において、各部分辞書パターンに対して用途に応じて認識に寄与する寄与係数を作成して保持しておき、前記照合手段による照合の際には、認識の用途に応じて寄与係数の分布から認識に有効な部分辞書パターンの集合を選択して使用するようにしたことを特徴とする画像認識方法。

40 【請求項26】請求項20の画像認識方法において、辞書データに基づき認識対象画像における各画素の輝度頻度を求めて保持する輝度頻度辞書を作成し、画像の認識時には入力画像と前記辞書とを照合するとともに、この照合の際には前記輝度頻度辞書と入力画像の各画素の輝度に基づき入力画像の各画素について信頼度を加味して照合を行なうことを有することを特徴とする画像認識方法。

【請求項27】人物の顔画像を含む対象の画像を取り込む取込工程と、  
前記取込工程にて取り込んだ対象画像に関し、この対象画像の所定領域内での輝度の分布の特異点に基づいて、特徴点を検出する特徴点検出工程と、

50 予め格納された認識のための特徴点の複数基準データに

(4)

5

関して照合のための有効性を示す複数寄与係数をそれぞれ算出し格納する寄与係数工程と、前記複数基準データの冗長性を考慮し、冗長な基準データを削除する基準データ削除工程と、前記基準データ削除工程にて冗長な基準データが削除された際の残りの複数基準データと、前記特徴点検出工程にて検出した特徴点とを、前記寄与係数工程にて格納している前記寄与係数に基づき、前記複数の単位領域ごとに比較し、前記対象画像を照合する照合工程と、を有することを特徴とする画像認識方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、対象の画像パターンを認識する画像認識装置であって、特に人間の顔を認識する人物顔認識装置とその人物顔認識方法に関するものである。

【0002】

【従来の技術】最近、入退出管理、ATM機器や券売機などの社会システムの個人照合、およびパソコン等OA機器のパスワードといった分野において、非接触でユーザに負担の少ない個人照合のニーズがあるが、その際にあまり特殊な装置を使用しないものとして画像認識の方法が注目されている。

【0003】これは被験者の顔画像を複数用意して、切り出し辞書とし、認識時には入力顔画像と辞書とのパターンマッチングを行って顔位置抽出と照合を行うものが一般的である。辞書としては、各個人でのさまざまな状況に対応させるために複数枚の画像を用意するものもあるが、そのままでは辞書サイズが膨大になるため、これらを統計的に特徴圧縮した辞書を作成しておく方式があり、統計的圧縮としては主成分の分析に類するものが使われている。

【0004】

【発明が解決しようとする課題】しかし上記した従来の画像認識による方法では、照明変動が少なくない環境では正確な認識が困難であるという問題がある。照明変動には、画面全体での輝度変化と、輝度勾配や影による部分的変化とがあり、前者に対して、辞書、および入力画像についてその画像ベクトルの大きさ（ノルム）による正規化を行ったり、ヒストグラム平坦化を行ったりした例があり、後者に対して、画面内輝度の平面当てはめをした例があるが、それらをもってしても顔抽出に失敗したり、照合が不正確になるという問題がある。

【0005】本発明は、上記問題を解決すべくなされるものであり、マッチングを行う領域の各小領域ごとの信頼度を計算し、この信頼度に応じたマッチングを行うことによって、迅速で正確なマッチングを実現し、これにより、人物の顔の迅速で正確な認識処理を行う人物顔認識装置とこの人物顔認識方法を提供することを目的とする。

6

【0006】

【課題を解決するための手段】本発明は、認識対象画像の認識を行なう画像認識装置において、認識対象画像を入力する画像入力手段と、この入力手段にて入力された認識対象画像から照合用の辞書を作成する辞書作成手段と、この作成手段にて作成した複数の辞書データに基づき認識対象画像における各画素の輝度頻度を求めて保持する輝度頻度辞書と、認識対象画像の認識時には入力画像と前記作成手段にて作成された辞書とを照合するとともに、この照合の際には前記輝度頻度辞書と入力画像の各画素の輝度に基づき入力画像の各画素について信頼度を加味して照合を行なう照合手段とを有することを有することを特徴とする画像認識装置である。

【0007】又本発明は、請求項1に記載の画像認識装置において、前記照合手段は前記輝度頻度辞書に基づき認識対象画像として尤もらしくない画素については照合に用いないようにしたことを特徴とする画像認識装置である。

【0008】又本発明は、請求項1に記載の画像認識装置において、前記照合手段は前記輝度頻度辞書に基づき認識対象画像として尤もらしくない画素については照合結果に対する寄与度を低くするようにしたことを特徴とする画像認識装置である。

【0009】又本発明は、請求項1に記載の画像認識装置において、前記輝度頻度辞書は前記作成手段による辞書データ作成時の切り出し画像の輝度頻度を求めて保持する顔領域抽出用輝度頻度辞書を有し、前記抽出手段は前記輝度頻度辞書に基づき認識対象画像の切り出しを行なう顔領域抽出手段を有することを特徴とする画像認識装置である。

【0010】又本発明は、請求項4に記載の画像認識装置において、前記辞書作成手段は、認識対象画像中の特定領域のパーツ画像に関するパーツ辞書を作成し、前記前記輝度頻度辞書は前記作成手段によるパーツ辞書作成時のパーツ画像の輝度頻度を求めて保持する顔パーツ抽出用輝度頻度辞書を有し、前記抽出手段は前記輝度頻度辞書に基づき前記顔領域抽出手段にて抽出された認識対象画像から顔パーツ画像の切り出しを行なう顔パーツ抽出手段を有することを特徴とする画像認識装置である。

【0011】又本発明は、請求項5に記載の画像認識装置において、前記辞書作成手段は認識対象画像中の特定領域の認識対象画像に関する顔照合辞書を作成し、前記前記輝度頻度辞書は前記作成手段による顔照合辞書の認識対象画像の輝度頻度を求めて保持する顔照合用輝度頻度辞書を有し、前記抽出手段は顔パーツ抽出手段にて抽出された顔パーツ画像との関係から入力画像中の特定領域の認識対象画像を照合用画像として選択する照合位置決定手段を有し、前記照合手段は顔照合辞書と照合用画像とを照合する際に前記顔照合用輝度頻度辞書に基づき基づき認識対象画像として尤もらしい画素を用いて照合



(5)

を行なうようにしたことを特徴とする画像認識装置である。

【0012】又本発明は、認識対象画像の認識を行なう画像認識方法において、入力された認識対象画像から照合用の辞書を作成し、この作成した複数の辞書データに基づき認識対象画像における各画素の輝度頻度を求めて保持しておき、認識対象画像の認識時には入力画像と前記辞書とを照合するとともに、この照合の際には前記輝度頻度辞書と入力画像の各画素の輝度に基づき入力画像の各画素について信頼度を加味して照合を行なうようにしたことを特徴とする画像認識方法である。

【0013】又本発明は、請求項7に記載の画像認識方法において、前記入力画像と前記辞書とを照合する際には前記輝度頻度辞書に基づき認識対象画像として尤もらしくない画素については照合に用いないようにしたことを特徴とする画像認識装置である。

【0014】又本発明は、請求項7に記載の画像認識方法において、前記入力画像と前記辞書とを照合する際には前記輝度頻度辞書に基づき認識対象画像として尤もらしくない画素については照合結果に対する寄与度を低くするようにしたことを特徴とする画像認識装置である。

【0015】又本発明は、認識対象画像の認識を行なう画像認識方法において、入力された画像から認識対象画像領域を切出し、切出された認識対象画像領域の画像中の特定領域の認識対象画像から照合用の顔照合辞書を作成し、かつ切出された認識対象画像領域における各画素の輝度頻度を求めて顔領域抽出用輝度頻度辞書を作成し、認識対象画像の認識時には入力画像中の画像のうち前記輝度頻度に基づき認識対象画像として尤もらしい領域を認識対象画像として抽出し、この抽出された認識対象画像中の特定領域の認識対象画像から照合用画像を選択して前記顔照合辞書と照合するようにしたことを特徴とする画像認識方法である。

【0016】又本発明は、請求項10に記載の画像認識装置において、前記顔照合辞書を作成する際に、認識対象画像中の特定領域のパーツ画像に関するパーツ辞書を作成するとともに、パーツ辞書作成時のパーツ画像の輝度頻度を求めてパーツ抽出用輝度頻度辞書を作成し、前記顔照合辞書と入力画像との照合を行なう場合には、前記顔領域抽出用輝度頻度辞書に基づき抽出された認識対象画像からパーツ抽出用輝度頻度辞書及びパーツ辞書に基づき入力画像から特定領域のパーツ画像を抽出し、抽出されたパーツ画像の位置関係から照合用画像を入力画像から選択して前記顔照合辞書と照合するようにしたことを特徴とする画像認識方法である。

【0017】又本発明は、請求項10に記載の画像認識方法において、前記顔照合辞書を作成する際に、顔照合辞書の認識対象画像の輝度頻度を求めて保持する顔照合用輝度頻度辞書を作成有し、前記顔照合辞書と照合用画像とを照合する際には前記顔照合用輝度頻度辞書に基づ

8

き認識対象画像として尤もらしい画素を用いて照合を行なうようにしたことを特徴とする画像認識方法である。

【0018】又本発明は、認識対象画像の認識を行なう画像認識装置において、認識対象画像を入力する画像入力手段と、この入力手段にて入力されたサンプル画像から特徴点を抽出する特徴点抽出手段と、この特徴点抽出手段にて抽出された各特徴点近傍の部分画像の特徴量を表す部分辞書パターンを形成して各特徴点の部分辞書を作成する辞書作成手段と、認識対象画像の認識時には前記入力手段にて入力された入力画像から前記特徴点抽出手段にて各特徴点を抽出して各特徴点近傍の部分画像から得られた特徴量と前記部分辞書の部分辞書パターンとの照合を行なう照合手段とを有することを特徴とする画像認識装置である。

【0019】又本発明は、請求項13の画像認識装置において、前記特徴点抽出手段は認識対象画像中の小領域内における輝度変化が特異な点又は輝度の幾何学的変化が特異な点を特徴点として抽出することを特徴とする画像認識装置である。

【0020】又本発明は、請求項13の画像認識装置において、前記辞書作成手段による辞書構築時に、サンプル画像から得られた複数の部分辞書パターンについての冗長性を評価し、非冗長とされた部分辞書パターンのみを部分辞書に保持することを特徴とする画像認識装置である。

【0021】又本発明は、請求項13の画像認識装置において、前記辞書作成手段による辞書構築時に、サンプル画像から得られた複数の部分辞書パターン金てを保持するのではなく、所定の間隔ごとに特徴量を示す部分辞書パターンを保持しておき、その間の区間については補完する近似関数を使用し特徴量を補完することを特徴とする画像認識装置である。

【0022】又本発明は、請求項13の画像認識装置において、各部分辞書パターンに対して用途に応じて認識に寄与する寄与係数を作成して保持しておき前記照合手段は寄与係数を加味して照合を行なうようにしたことを特徴とする画像認識装置である。

【0023】又本発明は、請求項13の画像認識装置において、各部分辞書パターンに対して用途に応じて認識に寄与する寄与係数を作成して保持しておき、前記照合手段による照合の際には、認識の用途に応じて寄与係数の分布から認識に有効な部分辞書パターンの集合を選択して使用するようにしたことを特徴とする画像認識装置である。

【0024】又本発明は、請求項13の画像認識装置において、辞書データに基づき認識対象画像における各画素の輝度頻度を求めて保持する輝度頻度辞書を有し、画像の認識時には入力画像と前記作成手段にて作成された辞書とを照合するとともに、この照合の際には前記輝度頻度辞書と入力画像の各画素の輝度に基づき入力画像の

(6)

9

各画素について信頼度を加味して照合を行なう照合手段とを有することを有することを特徴とする画像認識装置である。

【0025】又本発明は、入力された認識対象画像の認識を行なう画像認識方法において、入力されたサンプル画像から特徴点を抽出し、この抽出された各特徴点近傍の部分画像の特徴量を表す部分辞書パターンを形成して各特徴点の部分辞書を作成しておき、認識対象画像の認識時には入力された入力画像から前記特徴点を抽出して各特徴点近傍の部分画像から得られた特徴量と前記部分辞書の部分辞書パターンとの照合を行なうようにしたことを特徴とする画像認識方法。

【0026】又本発明は、請求項20の画像認識方法において、前記特徴点抽出においては認識対象画像中の小領域内における輝度変化が特異な点又は輝度の幾何学的変化が特異な点を特徴点として抽出することを特徴とする画像認識方法である。

【0027】又本発明は、請求項20の画像認識方法において、前記辞書作成時に、サンプル画像から得られた複数の部分辞書パターンについての冗長性を評価し、非冗長とされた部分辞書パターンのみを部分辞書に保持することを特徴とする画像認識方法である。

【0028】又本発明は、請求項20の画像認識方法において、前記辞書作成時に、サンプル画像から得られた複数の部分辞書パターン全てを保持するのではなく、所定の間隔ごとに特徴量を表す部分辞書パターンを保持しておき、その間の区間については補完する近似関数を使用し特徴量を補完することを特徴とする画像認識方法である。

【0029】又本発明は、請求項20の画像認識方法において、各部分辞書パターンに対して用途に応じて認識に寄与する寄与係数を作成して保持しておき前記照合手段は寄与係数を加味して照合を行なうようにしたことを特徴とする画像認識方法である。

【0030】又本発明は、請求項20の画像認識方法において、各部分辞書パターンに対して用途に応じて認識に寄与する寄与係数を作成して保持しておき、前記照合手段による照合の際には、認識の用途に応じて寄与係数の分布から認識に有効な部分辞書パターンの集合を選択して使用するようにしたことを特徴とする画像認識方法である。

【0031】又本発明は、請求項20の画像認識方法において、辞書データに基づき認識対象画像における各画素の輝度頻度を求めて保持する輝度頻度辞書を作成し、画像の認識時には入力画像と前記辞書とを照合するとともに、この照合の際には前記輝度頻度辞書と入力画像の各画素の輝度に基づき入力画像の各画素について信頼度を加味して照合を行なうことを有することを特徴とする画像認識方法である。

【0032】又本発明は、人物の顔画像を含む対象の画

10

像を取り込む取込工程と、前記取込工程にて取り込んだ対象画像に関し、この対象画像の所定領域内での輝度の分布の特異点に基づいて、特徴点を検出する特徴点検出工程と、予め格納された認識のための特徴点の複数基準データに関して照合のための有効性を示す複数寄与係数をそれぞれ算出し格納する寄与係数工程と、前記複数基準データの冗長性を考慮し、冗長な基準データを削除する基準データ削除工程と、前記基準データ削除工程にて冗長な基準データが削除された際の残りの複数基準データと、前記特徴点検出工程にて検出した特徴点とを、前記寄与係数工程にて格納している前記寄与係数に基づき、前記複数の単位領域ごとに比較し、前記対象画像を照合する照合工程とを有することを特徴とする画像認識方法である。

【0033】

【発明の実施の形態】以下、図面を用いて本発明を詳細に説明する。以下には、本方式を用いてビデオカメラから入力された画像中に含まれる人物の顔領域を切り出し、その姿勢を推定し、映像中の人物と辞書登録された人物との照合を行うための装置の二つの実施形態が示される。

【0034】(1)第1の実施形態の概要処理説明

図1は、本装置の実施形態の構成図である。図1において、この人物顔認識装置は、画像入力部1と、これに接続される画像蓄積部2と、これに接続されるユーザ操作部13と、これに接続される辞書作成部14と、これに接続される顔領域抽出用輝度頻度辞書7、顔パーツ抽出用輝度頻度辞書8、照合用輝度頻度辞書9、顔領域抽出部3、顔切り出し辞書10、顔パーツ抽出部4、顔パーツ辞書11、照合位置決定部5と、照合部6と、この照合部6に接続される顔照合辞書12と、照合部6に接続される表示部15とを有している。

【0035】この人物顔認識装置は、上述の顔切り出し、姿勢推定、顔照合の動作を行うために、複数撮影環境下の複数人物での顔画像サンプルから作成した顔切り出し辞書10、目、口などの顔パーツ辞書11を用いる。

【0036】図2は、本発明に係る第1実施形態に係る概略処理のフローチャートであり、以下、図2を用いて動作を説明する。本実施形態のシステムは、モノクロ、またはカラーのテレビカメラとA/D変換器とからなる画像入力部1により数フレームごとに入力されたデジタル画像データを一度、画像蓄積部2に格納する(S11)。そして、顔領域抽出部3においては顔領域抽出用輝度頻度辞書7に基づき、各画素について顔領域抽出用輝度信頼度を計算し(S13)、この顔領域抽出用輝度信頼度と顔切り出し辞書10とから、顔領域を抽出する(S15)。

【0037】更に、顔パーツ抽出部4においては顔領域内部の顔パーツ用輝度信頼度を計算し(S17)、この輝度信頼度と顔パーツ辞書とから、目、口を抽出する

(7)

11

(S19)。更に照合位置決定部5においては、顔パーツ領域の位置と入力画像とから照合用画像を生成する(S21)。そして最終的に、照合部6において照合用画像と登録済み辞書とのパターンマッチングを照合用信頼度辞書に従って算出する(S23)ことで、入力画像中の人物と辞書内登録人物との照合を行う。

【0038】つまり、顔領域抽出部3、顔パーツ抽出部4、照合部6においては、入力画像と顔辞書とのパターンマッチングを行うが、その際、本発明の第1実施形態の特徴である輝度信頼度を用いて、顔画像の辞書データのうち、輝度の分布の信頼度が一定水準であるものを中\*

$$Y = 0.30 \cdot R + 0.59 \cdot G + 0.11 \cdot B \quad (1)$$

以上が第1実施形態の概要であるが、以下に顔領域抽出部3、顔パーツ抽出部4、照合位置決定部5、照合部6の処理について詳細に説明する。

【0041】(a) 顔切り出し、照合辞書、顔パーツ辞書の作成

まず初めに、複数撮影環境、複数人物の顔のサンプル画像データを収集し、これらから顔が含まれる領域をマウスなどからなるユーザ操作部13によってマニュアルで切り出し、表示部15で確認した画像ベクトルを $I_i$

( $i=1, \dots, P$ ,  $P$ :全サンプル数)とし、これらのそれぞれを一定サイズ $N$ に正規化し、各画像内の照明変動を正規化した後、主成分分析して固有値の上位 $K$ 次元

( $k < N$ )を使用し、顔パターンを作成する。このとき生成された辞書としては、上記主成分分析による固有ベクトル( $N$ 次元)を固有値が大なる方から $k$ 個集めたものとなる。

【0042】ここでサイズについての正規化としては、正規化画像のサイズが一定となるように $I_i$ を一定割合で間引いたものとして良い。また照明変動に対する正規化としては、本明細書の“従来の問題点”で述べた“画像ベクトルの大きさによる正規化”(ベクトルの各成分をベクトルのノルムで割っておく)、“ヒストグラム平坦化”[1]、“画面内輝度の平面当てはめ”[1]などを使用して良い。

【0043】以上において、辞書作成用サンプル画像データとして複数撮影環境、かつ各登録者 $j$ ごと(ただし $j=1, \dots, P_j$ ;  $\Sigma P_j = P$ )のデータのみを使用して作成したものを顔照合用辞書12とし、複数撮影環境、かつ複数人のデータを全て使用して作成したものを顔切\*

$$1/|V| \quad \Sigma \quad (V, Dk)$$

$k=1, K$

ただし、 $V$ :切り出し部分画像サイズ正規化画像ベクトル、 $Dk$ :顔切り出し辞書の第 $i$ 固有ベクトルであり、★

$$(V, Dk) = \Sigma v(i) \times dk(i) \quad (3)$$

$i=1, M$ となる。ただし、 $v(i)$ ,  $dk(i)$ はそれぞれのベクトルの第 $i$ 要素。

【0048】また距離による方法では、上述の $V$ と辞書サンプルベクトルの代表値たとえば上記 $D1$ とを $Dk$

12

\*心にマッチングを行うことで、迅速にしかも一定の認識精度を確保する人物認識装置を実現するものである。

【0039】なお、上記において画像入力部1がモノクロカメラからの入力を対象とした場合には、画像蓄積部2は、8bitsのモノクロ階調画像として蓄積し、カラー画像を対象とした場合には入力画像(RGB)の色成分を輝度成分 $Y$ に変換して蓄積する。このときの変換式は一般に用いられている以下のようなもので差し支えない。

【0040】

※り出し辞書10とする。

【0044】顔パーツ辞書11も同ようにして作成されるものであるが、サンプル画像としては、顔全体を用いるのではなく、目、口の領域のみを使用する。

(b) 顔領域抽出、顔パーツ抽出

次に、顔領域抽出部3で行う入力画面内の顔位置の検出について、図2に示すフローチャートに従って説明する。顔領域抽出部3では、画像蓄積部2に一時的に蓄積された画像(原画)について、顔位置の検出が行われる。

【0045】初めに、輝度頻度辞書7から顔領域抽出用の輝度分布を取得する(S31)。そして、原画のある解像度に縮小した縮小画像を作成する(S33)。次に、縮小画像の画面ごとの顔領域抽出用輝度信頼度を算出(更新)する(S35)。そして、この縮小画像内をラスタスキャンして、前述の顔切り出し辞書11内に蓄積された特徴ベクトルサイズにしたがって、部分画像を切り出し、特徴ベクトルとの間のマッチングをとる(S37)。更に、そして縮小率を複数変えた場合において、前述のマッチングの評価がもっとも良いときの位置を顔領域とする(S39、S41)。

【0046】マッチング手法としては、文字や音声のパターン認識手法として一般的なように、辞書内特徴ベクトルと部分画像ベクトルとについての、内積最大なるものを検出するか、距離(例えばユークリッド距離、マハラノビス汎距離等)最小なるものを検出するという方法を使用して差し支えない。

【0047】具体的には内積による方法では、

(2)

★ $|V|$ は $V$ のノルムである。 $(V, Dk)$ は $V$ と $Dk$ の内積であるが、両者の次元数を $M$ 次元とすると

( $k=1, \dots, K$ )に射影してそれぞれ射影ベクトル $P_V, PD1$ を作成し、それらのユークリッド距離1を求める。

【0049】ここに

(8)

13

$$PV(k) = (V, Dk)$$

ただし、 $k=1, \dots, K$ であり、 $PD1(k) = (D1, Dk)$ である。

【0050】以上で式(3)、式(3)'においてVの各要素は入力画像内から切り出された部分画像を一定サイズに正規化して作成した画像の各画素に相当するが、これらについて後述の輝度頻度辞書から画素ごとに判定\*

$$(V, Dk) = PV(k) = \sum$$

$i=1, M$

ここに $T(i)$ は上記信頼度である。上記で各解像度画像が生成されるごとに輝度信頼度も後述の方法で更新する。

【0052】顔パーツ抽出部4ではパーツとして目、口を抽出するが、入力画像中で、上記顔領域抽出部3で抽出された顔領域内部に対して、これと同ようなパターンマッチングを顔パーツ辞書11を用いて行う。図4は、顔パーツの抽出処理を示すフローチャートである。図4において、注目パーツとして左目、右目、口のうちの一種類を設定する(S51)。このとき、目、口というパーツごとに、注目パーツ用の別々の輝度頻度を使用して信頼度を算出する(S53)。そして、顔領域内部で輝度信頼度と注目パーツ辞書とから、注目パーツを抽出する(S55)。このようにして全パーツを抽出するまで(S57)、注目パーツを次々と替えていき、注目パーツの抽出を続ける(S59)。

【0053】(c) 輝度頻度辞書の作成と信頼度判定次に、輝度頻度辞書(7, 8, 9)の作成とその信頼度\*

$$T(x, y) = \{1/\sum h(x, y, I)\} \times h(x, y, I)$$

$$I = I_{\min}, \dots, I_{\max}$$

$$= 0$$

ただし上記で $\alpha$ は信頼度の範囲を決定するための任意の定数であり、 $I_{\max}$ ,  $I_{\min}$ は観測画像の輝度値の最大値と最小値である。

【0058】また別手法としては文献[2]に示すような事後確率変換による手法を用いても良い。ここで、事象 $\theta_0$ を信頼できるもの、事象 $\theta_1$ を信頼できないもの★

$$w1' = \{w1 p(I|\theta1)\} / \{p(I|\theta0)(1-w1) + p(I|\theta1)w1\} \quad (5)$$

となり、この $w1'$ を信頼度として用いる。

【0059】ここで、顔辞書作成時の切り出し画像の頻☆

$$p(I|\theta0) = \{1/\sum h(x, y, I)\} \times h(x, y, I) \quad (6)$$

$I = I_{\min}, \dots, I_{\max}$ であり、 $p(I|\theta1)$ は一般にデ ◆ ◆一タサンプルが無いため、

$$p(I|\theta1) = 1/(I_{\max} - I_{\min}) \quad (7)$$

と推定する。また上記は入力画像が更新されるたびに $w1 = w1'$ と更新し、 $w1$ の初期値 $= 1.0/2.0$ としてよい。

【0060】(d) 照合位置の決定

照合位置決定部5では、顔領域抽出部3で求められた顔領域と顔パーツ抽出部4で求められた顔パーツとの関係

14

(3)'

\*した信頼度を用いて重み付けを行い、輝度として尤もらしい画素のみを使用してマッチングを行うのが本実施形態のポイントである。したがって、(3)、(3)'は以下ようになる。

【0051】

$$v(i) \times T(i) \times dk(i)$$

(3)''

10 ※の判定について、以下に述べる。

【0054】前述の(a)のように、複数撮影環境における複数登録者についてのサンプル画像を収集し、顔領域として切り出された画像 $I_i$  ( $i=1, \dots, P$ ,  $P$ :サンプル数)について一定サイズ $N$ に正規化した画像を $J_i$  ( $i=1, \dots, P$ ,  $P$ :サンプル数)とする。

【0055】まず、各 $J_i$ の各画素点 $(x, y)$ ごとに全サンプルに対する輝度の頻度分布 $h(x, y, I)$ を作成し、輝度頻度辞書(7, 8, 9)とする。そして認識時には $h(x, y, I)$ と入力部分画像の対応画素 $I(x, y)$ とから信頼度 $w(x, y)$ を算出する。ここで輝度頻度辞書としては、顔抽出用7、顔パーツ抽出用8、顔照合用9として別々のものを以下のように準備する。

【0056】上記のような信頼度の算法としては様々なものが考えられるが、たとえば輝度分布 $h(x, y, I)$ の平均 $Mh$ と標準偏差 $\sigma h$ とから信頼度 $T(x, y)$ を以下のように決定しても良い。

【0057】

$$\text{if } |I - Mh| < \alpha \times \sigma$$

else (4)

★とする。ある部分画像内の画素の輝度を観測する前の $\theta_0$ 、 $\theta_1$ の事前確率がそれぞれ $w_0$ 、 $w_1$ であるとき(ただし $w_0 + w_1 = 1$ )、事後確率 $w1'$ は事象 $\theta_k$ のもとでの画素輝度 $I$ の条件付き確率を $P(I|\theta_k)$ とするときベイズの定理から

10 ☆度分布 $h(x, y, I)$ から

から、人の顔として経時変化の少ない領域として顔の目と口の端が結ぶ四辺形領域を選択し、これが照合辞書12での顔パーツの位置に一致するようにアフィン変換や[3]に示す幾何変換を用いて変形した切り出し領域からなる照合用画像を作成する。

50 【0061】(e) 照合

(9)

15

最後に、図5のフローチャートを用いて照合の手順を説明する。照合位置決定部5によって作成された照合用画像について、照合辞書12と(b)で示したのと同じの手法によってパターンマッチングを行うが、この際には、照合対象としては、照合辞書12内の登録者jごとの辞書Djとなる。

【0062】つまり、注目登録済み人物の辞書として、個人iのものDiを設定する(S61)。そして、注目人物の輝度信頼度を計算する(S63)。更に、照合領域内部で照度信頼度とDiとから注目人物との類似度(距離)を算出する(S65)。ここにおいても、切り出し画像内での各画素について判定した信頼度を用いた(3)〃式に基づいたマッチングを行う。ここで、照合用画像と辞書Djとのマッチングを行うにあたり、輝度頻度分布は個人辞書ごとに異なるため、照合用辞書ごとに異なる分布を使用してもよいし、登録者全体の頻度分布を1つ作成し、これのみを使用してもよい。

【0063】更にこのマッチングを全登録者と照合するまでこれを行い(S67、S71)、最適なマッチングが行われた人物のIDを出力することで、照合が終了する(S69)。

【0064】更に、このマッチングの際の信頼度の扱い方であるが、その方法が少なくとも二つ考えられる。一つは、信頼度の程度に応じて、この値を係数としこの大きさに応じてマッチングを行う。つまり、辞書データのなかで、信頼度の高い単位領域(メッシュ)については、原画像と100%に近いマッチングを行う。一方、信頼度が低い単位領域(メッシュ)については、低い程度にマッチングの程度を減らしていくものである。こうすることにより、輝度の程度に応じた、最適なマッチングを実現することができる。

【0065】もう一つは、信頼度に関するしきい値を設け、この信頼度が一定値以下の値であれば、この信頼度に関する辞書データの単位領域(メッシュ)については、原画像とのマッチングを行わない。一定値以上であれば、この辞書データの単位領域(メッシュ)は原画像とのマッチングを行うというものである。

【0066】(2)第2の実施形態の概要処理説明次に、本発明の第2の実施形態について、以下に図面を用いて詳細に説明する。

【0067】第1の実施形態の特徴は、辞書データの輝度に関する信頼度を求め、この信頼度に応じて辞書データを扱うことで、本質的な識別精度を落とすことなく、マッチングの頻度を減らしていくことで、認識処理の速度を向上させるものである。

【0068】一方、第2の実施形態においては、本発明の第2のポイントとして、辞書データの各単位領域(メッシュ)の内、識別処理に寄与する程度を表す寄与係数なるものを計算し、これを各辞書データの各単位領域に対応させてあらかじめ記憶させておく。そして、寄与係

16

数の低い単位領域については、例えばこれは、画面上の顔領域以外の、例えば背景の画像がこれに対応するが、原画像とのマッチングの際に、マッチングの対象としないようにして、マッチングに係る時間を短縮化し、認識処理の迅速化を図るものである。

【0069】図6は、第2の実施形態の構成図である。図6において、この人物顔認識装置は、画像入力部1と、これに接続される画像蓄積部2と、これに接続される特徴点抽出部18と、これに接続される辞書作成部19と、これに接続される辞書再評価部20と、これに接続される部分特徴辞書23と、これに接続される顔領域抽出部21と、これに接続される顔姿勢推定部22と、これに接続される照合部6と、辞書作成部19に接続される顔領域抽出用寄与係数24、顔姿勢推定用寄与係数25、照合用寄与係数26と、更に辞書作成部19に接続される顔領域抽出用輝度頻度辞書27、顔姿勢推定用輝度頻度辞書28、照合用輝度頻度辞書29と、照合部60に接続される表示部31と、これに接続されるユーザ操作部30とをそれぞれ有する。

【0070】このような構造において、第2の実施形態の人物顔認識装置は、顔領域抽出、顔姿勢推定、照合の各用途ごとに各辞書内射影データの認識有効性を示す寄与係数が辞書作成部19により作成される。更に、この寄与係数の程度に応じて、辞書データの単位領域(メッシュ)が、識別のためのマッチングに使用されたり、使用されなかったりして、実質的な識別精度を保持しながら、照合処理の迅速化を図るものである。

【0071】以下、第2の実施形態の処理の動作と、その他の本発明のポイントを、図7、図8のフローチャートを用いて説明する。本実施形態のシステムは、モノクロ、またはカラーのテレビカメラとA/D変換器とからなる画像入力部16により数フレームごとに入力されたデジタル画像データを一度、画像蓄積部17に格納し(S71)、特徴点抽出手段18で得られた特徴点周囲の画像を切り出し、辞書作成手段19によって前述の辞書23、27、28、29を作成する。

【0072】まず、画像入力部16により数フレームごとに入力されたデジタル画像を入力し画像蓄積部17に格納する。更に、本発明の第3ポイントとして、顔切り出し、姿勢推定、顔照合の動作を行うために、特徴点抽出部18により、複数撮影環境下の複数人物での顔画像サンプルから認識に有効な特徴点の抽出を効率的に行なう(S73)。そして、辞書作成部19においては、特徴点の近傍の小領域を窓としてもち、それら窓画像を主成分分析した部分空間と、各窓をその空間へ射影した射影ベクトルとを辞書24として作成する(S75)。この方式によれば、入力画像に対しても同ような窓領域内画像の前記部分空間への射影を計算し、それらと辞書内射影とのマッチングを用いることで、認識時の顔画像の部分的照明変動や他の物体による顔領域の一部の隠れに

(10)

17

対しても、誤認識の発生を抑制することができる。

【0073】更に本発明の第4ポイントとして、辞書再評価部20において辞書内射影データの冗長性の評価と圧縮を行う。つまり、辞書データ内に射影データが繰り返し同様のパターンを描いて存在している場合、原画像とのマッチングを行う際に、比較の意味があまりないようなデータ、冗長性の高いデータを削除することで、辞書データを削減し、これによりマッチング速度を向上させていく(S77)。

【0074】次に本発明の第5ポイントとして上げられる、各サンプル内部の画像射影変動の近似関数表現を行い、これにより、辞書データの効率的な削減を行う(S79)。つまり、辞書データが、近似関数により表すことができる場合、これを近似関数に表すことで、データ量を大幅に削減することができる。これにより、マッチング速度を非常に向上させることができる。

【0075】このような画像サンプルをマニュアルで用途ごとに分類する(S81)。そして、辞書作成部19においては、初めに上げた本発明の第2ポイントとしての特徴である、寄与係数の算出を行う。つまり、辞書データの各単位領域(メッシュ)の内、識別処理に關与する程度を表す寄与係数を計算し、これを各辞書データの各単位領域に対応させてあらかじめ記憶させておく(S83)。そして、寄与係数の低い単位領域については、例えばこれは、画面上の顔領域以外の、例えば背景の画像がこれに対応するが、原画像とのマッチングの際に、マッチングの対象としないようにして、マッチングに係る時間を短縮化し、認識処理の迅速化を図るものである。

【0076】そして、本発明の第6ポイントとして、全辞書サンプルに対する統計分布からしきい値以上に寄与係数の高いもののみを選択してマッチングに使用することで、処理速度向上を図るものである(S85)。

【0077】次に入力画像に対して、特徴点抽出部18で抽出された特徴点周囲の部分画像と辞書内射影データとのマッチングの結果から、顔領域推定部21、顔姿勢推定部22、照合部23において、顔が画面内のどこに存在し、どういう向きで、だれの顔であるかを求める。これらの各処理においては前述の辞書内の寄与係数にしたがった重み付をおこなったマッチングを用いる。

【0078】又、以上のマッチングにおいて、更に、本発明の第7ポイントとして、第1の実施形態と同ような手法を適用して、辞書作成部19により輝度頻度を算出し輝度頻度辞書(27, 28, 29)に記憶しておく。そして、輝度頻度分布に基づいた信頼度を算出して利用することにより、辞書データの非常に洗練されたマッチングが実現され、認識精度を保持したまま認識速度のものが向上することとなる。

【0079】なお、第1の実施形態と同ように、画像入力部1がモノクロカメラからの入力を対象とした場合に

18

は、画像蓄積部2は、8bitsのモノクロ階調画像として蓄積し、カラー画像を対象とした場合には入力画像(RGB)の色成分を(1)式にしたがって、輝度成分Yに変換して蓄積する。

【0080】次に図8は、顔画像の認識を行う際の動作を説明したフローチャートであり、まず入力部16により認識たいそう画像である顔画像を画像蓄積部17に記憶し(S91)、次に特徴点抽出部18において、入力顔画像内の特徴点を抽出して、その周囲の部分画像を切り出す(S93)。更に、顔領域抽出部においては切り出された部分画像と辞書データとの間のマッチングが行われ、顔領域が推定される(S95)。この際には、顔領域抽出用寄与係数24及び顔領域抽出用輝度頻度辞書27に基づく輝度頻度を併用して顔領域が推定される。更に顔姿勢推定用係数25と顔姿勢推定用輝度頻度辞書28に基づき算出された輝度信頼度とに基づく部分画像と辞書23とのマッチングが行われ、顔姿勢が推定される(S97)。

【0081】さらに照合部60においては、照合用寄与係数26と照合用輝度頻度辞書に基づき算出された輝度信頼度とに基づき部分画像と辞書23とのマッチングが行われて、顔画像がだれであったのかが判定される(S99)。

【0082】次に、上述した処理のうち「特徴点の抽出」、「辞書の構築と圧縮」、「マッチング方法」の細部を説明する。以上、第2の実施形態の処理動作の概要を説明した。

【0083】(a) 入力画像からの特徴点抽出

更に、本発明の第3ポイントである、画像の特徴部分を抽出することにより、画像認識を行う工程について、以下に詳細に説明する。特徴点抽出部18は本発明の第3ポイントとして、以下の方法で画像内から認識に有効な特徴を抽出する。

【0084】入力画像と辞書画像のマッチング、および連続フレーム間でのマッチングなどを行う際に、ある領域を選択してその周囲をテンプレートとすることが一般的である。この理由の1つは高速化のためであるが、もう1つは、対象物に相当する全画素を追跡処理対象とした場合に比べ、よりマッチング(主として距離や相関による)に有効な点、すなわち他の点と区別が付くような点を選択するためである。このようなものとして画像内のエッジが良く知られている。

【0085】ここではそのようなテクスチャ特異な有効特徴を求めるために[4]の有効特徴点抽出手法を用い、これに前処理を施すことで高速化する。辞書作成用画像や認識用入力画像全面について、[4]の処理を行う場合、各画素点についての2×2行列の固有値を求める必要があり、処理コストがかかる。そこで、あらかじめ全処理領域内テクスチャが幾何学的に特徴的である点として後述の方法でコーナー点を抽出し、有効特徴点抽

(11)

19

出領域をコーナ一点近傍のみに限定する。

【0086】以下の処理の工程を図9に基づき説明する。ここにおいてのコーナ検出は経験的に良い特徴量だとされるエッジを5×5画素のマスクサイズ内において評価してコーナ点を判定する。

【0087】まず、画像入力部16により画像を入力したら(S101)、あらかじめ原画にsobel[5]\*

$p \geq 2 \times \text{最大画素値} \rightarrow \text{"+"型分岐点}$

$p \leq -2 \times \text{最大画素値} \rightarrow \text{"x"型分岐点}$

上記以外  $\rightarrow$  コーナ候補ではない

次に"+"型、"x"型それぞれについて、図10

(b)のマスクによる場合分けで上記と同ようにして、※

−最大画素値  $\leq p < \text{最大画素値} \rightarrow$  コーナ点

上記以外

"x"型分岐点の場合は

−最大画素値  $\leq p < \text{最大画素値} \rightarrow$  コーナ点

上記以外

と場合分けする。

【0089】最後にコーナと判定された注目点p(x, y)の周囲領域W内について次式にしたがって特徴点を★

$$G = \sum_{W} (gx^* gx \quad gx^* gy) = (a \quad b) \quad (8)$$

$$W(gy^* gx \quad gy^* gy) \quad (c \quad d)$$

次にGの固有値 $\lambda+$ 、 $\lambda-$ を求め、そのうちの大きい方を注目点の画素値とする。上記処理を処理領域の全画素☆

$$\lambda > \lambda_t$$

なる画素値の部分の有効特徴点として抽出する。このときの閾値は、処理領域内部の各画素値について判別分析によって求めたものを使用する。

【0091】(b)辞書の構築と圧縮

辞書作成部19では、各顔画像サンプルについて、前節(2)(a)の方法で抽出された認識有効特徴点の周囲W×H画素の領域内部の部分画像を切り出し、それらを主成分分析し、その内の第K固有ベクトルまでを求めた行列(辞書空間行列)と、その空間への全顔サンプルについての全部分画像ベクトルの射影とを求めて、辞書とする。

【0092】しかし、サンプル数が多い場合には上記辞書は膨大になるため、上記した本発明の第4ポイント、◆

$$D_{ij} = 1/j \cdot \sum_{k=1}^K |P_i - P_j|$$

$$j = 1, \dots, K$$

ただし、 $P_j$ ：画像J内の部分画像のうち、画像I内の部分画像と射影距離が最小となるもの、K：その数である。また距離としてはユークリッド距離や絶対値距離を用いて良い。

【0094】以上を全サンプルについて計算した結果のテーブルの例を図11に示す。これらの距離(図中のD12~Dnn-1)の頻度分布を作成し、頻度分布の分散比を最大にするように判別分析法[6]を適用して閾値を求める。このとき閾値より距離が近かったサンプル画像内の部分画像については、その射影を辞書から除外する。

20

\*をかけ、判別分析2値化[6]を行う(S103)。そして、この画像中の各画素について、注目画素が1の場合に図10(a)のマスクによるconvolutionをとる。そして結果の画素値pの値によって、注目点が"+"型分岐点に近いか"x"型分岐点に近いかを判定する。ここでは、以下のように場合分けした。

【0088】

※コーナ点を求める(S105)。すなわち"+"型のときは以下のように場合分けし、

$\rightarrow$  コーナ点

$\rightarrow$  コーナ点でない

★判定する(S107)。まず以下のGradient matrix Gを求める。

【0090】

☆について行った後、ある閾値 $\lambda_t$ について

(9)

◆第5ポイントに従って、辞書容量を削減していく。まず、本発明の第4ポイントとしての方法では、あるサンプル画像I内部の各部分画像 $W_i$  ( $i=1, \dots, NI$ )の辞書空間への射影 $P_i$ を求め、他サンプル画像J ( $J \neq I$ )内の部分画像のうち、その射影と $P_i$ との距離が最小なるもの $w_j$ とその距離 $P_j$ を求める。これによって、 $w_j$ の属するサンプル画像が求められるため、画像I内の部分画像 $W_i$ とマッチする他サンプル画像の頻度分布が得られる。このとき、頻度分布最大なる他サンプル画像 $J_{max}$ が画像Iとマッチしたと定義する。またこのときのIと $J_{max}$ との距離 $D_{ij}$ を以下で定義する。

【0093】

(10)

【0095】また本発明の第5ポイントとしての方法では、第3項の方法でサンプル画像単位で辞書削減を行った後に適用するものとする。まず、辞書作成用の顔画像サンプルをある個人について、あるパラメータ例えば照明や方向の変動の順番にソートする。このようにすると一般にとなりあうサンプル画像間内の部分画像の射影同士も連続する。

【0096】次に上述のサンプル画像の全部分画像について、各部分画像間での射影距離を求めるが、もし、ある部分画像 $w_i$ からみたときに他の部分画像 $w_j$ が最短距離にあった場合に、同時に $w_j$ からの最短距離部分画



(12)

21

像が $w_i$ であるときには、両者を類似と見なす。そして、隣接するサンプル画像列中の部分画像 $w_{i1}$ ,  $w_{i2}$ ,  $w_{i3}$ , ...  $w_{in}$ が類似していた場合、これらをまとめて近似関数で表現し、辞書サイズを削減する。近似関数としては、ようなものが考えられるが〔7〕などに用いられているキュービックスプラインなどを用いて良い。

【0097】さらに辞書作成部19では、本発明の第2ポイントとして、各部分辞書パターンについて、顔抽出姿勢推定、照合の用途に応じて、有効な認識のための寄与係数を辞書サンプルから作成し、認識時に各用途ごと\*

$$\sum P_i \times V_i$$

$i = 1, k$ により用途ごとの重み画像を作成する。

【0099】最後に、各サンプル画像 $I$ を上記重み画像に重ね合わせたときの各部分画像 $w_i$ 領域内の重み画素\*

$$C_{wi} = S / P_{max}$$

ここに $P_{max}$ は上記重み画像内の最大画素値である。なお用途ごとのサンプルとしては、例えば以下のようにする。

【0101】顔領域推定の場合：各サンプル画像の特徴画像で顔領域外の値を0としたもの

顔姿勢推定：同上

顔照合：個人ごとのサンプル画像集合について、特徴画像を作成したもの

この場合には個人辞書ごとに個別寄与係数を作成

また上記では、用途ごとの寄与係数によって、用途に応★

$$C_{wi} < C_T$$

なる部分画像辞書はマッチングに使用しない。

【0102】(c) マッチング方法

顔領域推定部21、顔姿勢推定部22、照合部23、においては、入力画像内について以下の手順でマッチング30

【0103】工程1：入力画像内から部分画像を抽出し、それらの各々について、辞書内に登録された部分画像辞書うち類似度最小なるものを求める(マッチング)。

工程2：上記のマッチング部分辞書パターンの含まれるサンプル用画像のうち、最も出現頻度の高いもの $I$ とそのときの頻度 $H$ を求める。

【0104】工程3：信頼度 $T = (\text{最大出現頻度 } H / \text{入力画像内の全局所ウインドウ数})$ を求める。

工程4：上記 $I$ に相当する局所ウインドウについて、辞書から座標のシフトをそれぞれ求め、その投票演算から☆

$$P_i(k) = \sum_j I(j) \times C(i) \times d_k(j) \quad (14)$$

$j = 1, M$

ただし、 $M$ は部分画像ベクトルの次元数、 $C(i)$ は部分画像 $I_i$ の寄与係数、 $d_k(j)$ は辞書空間の第 $k$ 固有ベクトルの $j$ 番要素である。

【0108】さらに本発明の第7ポイントとしての輝度◆

$$P_i(k) = \sum_j I(j) \times C(i) \times d_k(j) \times T(i) \quad (15)$$

$j = 1, M$ のようにすることで、用途ごとの認識寄与に30

22

\*に部分辞書パターンの認識結果への影響の仕方を調整可能とする。以下に、寄与係数の算出方法の例をあげる。

【0098】まず用途ごとにそろえた複数サンプルについての、前述の特徴点画像を求める。この画像の輝度分布がサンプル画像内で特異性を示すため、最適認識のためには、その特徴点近傍部分画像における、その用途での認識寄与は高くあるべきである。そこで、前記特徴点画像を用途サンプルごとに主成分分析して、上位 $K$ 個の固有値 $P_i$ 、固有ベクトル $V_i$  ( $i = 1 \dots n$ )を抽出し

(11)

※値の総和 $S$ によって以下のように各部分画像ごとの寄与係数 $C_{wi}$ を算出する。

【0100】

(12)

★じて、冗長性の高い部分画像辞書の認識処理への影響をなるべく少なくし、精度を向上するのが目的であるが、さらに寄与の低い部分画像辞書でのマッチングによる処理時間増大を防ぐためには、本発明の第6ポイントとして既に述べたように、これら係数の分布から閾値を統計的に決定し、その閾値以下の寄与係数の部分画像辞書は使用しなければ良い。このために、例えば、上述の部分画像辞書ごとに求められた寄与係数の頻度分布をとり、判別分析法〔6〕により閾値 $C_T$ を求め、各部分画像辞書ごとに与えられた寄与係数 $C_{wi}$ について

(13)

☆シフト量の頻度最大なるウインドウを選びそれらと辞書からのアフィン変換を求める。

【0105】上記において、工程1、工程2、工程3の処理を照合部23が行い、また、工程1～工程4を顔姿勢推定部22が行う。最後に顔領域推定部21において、工程1～工程4に続いて、サンプル画像 $I$ での顔画像領域に対して、上述のアフィン変換を施し、入力画像内での顔領域を推定する。

【0106】なお、上記での $I$ を推定値とし、推定信頼度 $T$ とアフィン変換係数によってリジェクトを行う。以上で工程1においては、全節で述べた用途ごとの寄与係数 $C$ を用い、(3)と同様に部分画像 $I_i$ の辞書空間への射影 $P_i(k)$ を求めると寄与係数を考慮したマッチングが行える。

【0107】

◆の信頼度を(1)(c)と同様に求め、認識時に入力画像から、上記と同様に部分画像を切り出し、その領域内の各座標 $(x, y)$ に対して(4)式の $T(x, y)$ や(5)式の $w_{1'}$ のように求め、それを $T(i)$ とすると、

加え、輝度の信頼度を考慮したロバストな認識が可能と



(13)

23

なる。

【0109】以上、本発明の人物認識装置によれば、以下の効果が考えられる。つまり、入退出管理、ATM機器や券売機などの社会システムでの個人照合、およびパソコン等OA機器のパスワードといった分野において、非接触でユーザに負担の少ない個人照合部のニーズがあるが、その際にあまり特殊な装置を使用しないものとして画像認識の方法が提案されているが従来装置では画像認識では照明変動が少なくない環境では正確な認識が困難であるという欠点がある。一方、本発明によれば、辞書作成サンプル画像内の輝度分布の信頼性を入力画像内の各画素輝度について検証し、輝度について信頼性の高い画素、またはその周辺の画素のみを認識に利用することで、輝度の大きな変化があっても安定な認識を行うことを可能とする。

【0110】また、顔辞書として顔のある領域全体の画像とするのではなく、顔画像に含まれる部分特徴領域近傍の部分領域を用いて作成し（部分辞書パターン）、それらのマッチング結果を総合判定することによって多少の輝度変動や、他物体による隠れなどによる変動を吸収することができる。

【0111】この際、前記部分辞書パターンのデータサイズが大きくなって、本手法を実現するための装置規模が大きくなることが問題となる。しかし、辞書作成サンプル画像ごとの冗長性を統計的に評価して冗長なパターンを除去したり、特徴量変化が連続するようなサンプル間においては、各部分辞書パターンも、それ自身ではなく補完関数で近似して保持するなどにより辞書サイズを小さくできる。

【0112】さらに顔抽出、姿勢推定、照合という各処理に応じて、最適な部分パターンの寄与を考慮して使用することにより、認識の精度と処理時間との両面から顔認識性能を著しく向上することが可能となる。

【0113】なお、参考文献を以下に示す。

[1] K. Sung and T. Poggio., Example-based Learning for View-based Human Face Detection. In Proceedings from Image Understanding Workshop, pp. 843-850, November 1994.

[2] 中井：事後確率を用いた移動物体検出手法，情報処理，94-CV-90, pp. 1-8, 1994

[3] 山口、福井、前田、動画像を用いた顔認識システム、信学技報、PRMU97-50, pp17-23, (1997)

[4] Shi, Tomasi, Good Features to Track, IEEE Conference on Computer Vision and Pattern Recognition (CVPR94) Seattle, June 1994

[5] 高木、下田、“画像解析ハンドブック”、東京大学出版会、pp. 553-554、1991

[6] 大津、判別および最小二乗基準に基づく自動しきい値選定法、信学論、Vol. J63-D, No. 4, pp. 349-356, 1980

[6] 高木、下田：画像解析ハンドブック PP. 578

24

[7] 村瀬、ネイヤー、“多重解像度と固有空間表現による3次元物体のイメージスポッティング、情報処理、Vol. 36 No. 10, pp2234-2243, Oct, 1995

【0114】

【発明の効果】以上、詳細に説明したとおり、本発明によれば、顔画像の認識処理を行うための辞書データに対して輝度に関する信頼度を求め、これに関連づけて格納しておくことで、認識処理の際に、例えば辞書データが作成された際の照明が不適当な場合の認識にあまり適さない辞書データの単位領域（メッシュ）に対しては、カメラからの原画像とのマッチング処理に用いず、又は用いても低い程度でのマッチング処理を行う。これにより、確実な認識を実現するメッシュに関してはマッチングを確実にし、信頼度の低いメッシュは省略することで、全体として、識別精度を保持しながら、迅速な人物の顔画像の識別処理を実現することができる。

【0115】又更に、辞書データの単位領域（メッシュ）それぞれにつき、マッチング処理にどれ程寄与するかという観点から、寄与係数をそれぞれ設け、マッチングの際には、この寄与係数に応じて有効な複数メッシュを選択的に基準としてマッチングを行う。これにより、識別精度を保持しながら、辞書データを削減することによって、迅速な人物の顔画像の識別処理を実現することができる。

【図面の簡単な説明】

【図1】本方法の第1実施形態に係る全体構成図。

【図2】本発明の第1実施形態に係る概略処理のフローチャート。

【図3】本発明の第1実施形態に係る顔領域抽出のフローチャート。

【図4】本発明の第1実施形態に係る顔パーツの抽出のフローチャート。

【図5】本発明の第1実施形態に係る顔照合処理のフローチャート。

【図6】本方法の第2実施形態に係る全体構成図。

【図7】本発明の第2実施形態での辞書作成のフローチャート。

【図8】本発明の第2実施形態での認識のフローチャート。

【図9】本発明の第2実施形態に係る特徴点抽出手法。

【図10】本発明の第2実施形態に係るコーナー抽出マスク。

【図11】本発明の第2実施形態に係るサンプル画像間の距離例。

【符号の説明】

1…画像入力部

2…画像蓄積部

3…顔領域抽出部

4…顔パーツ抽出部

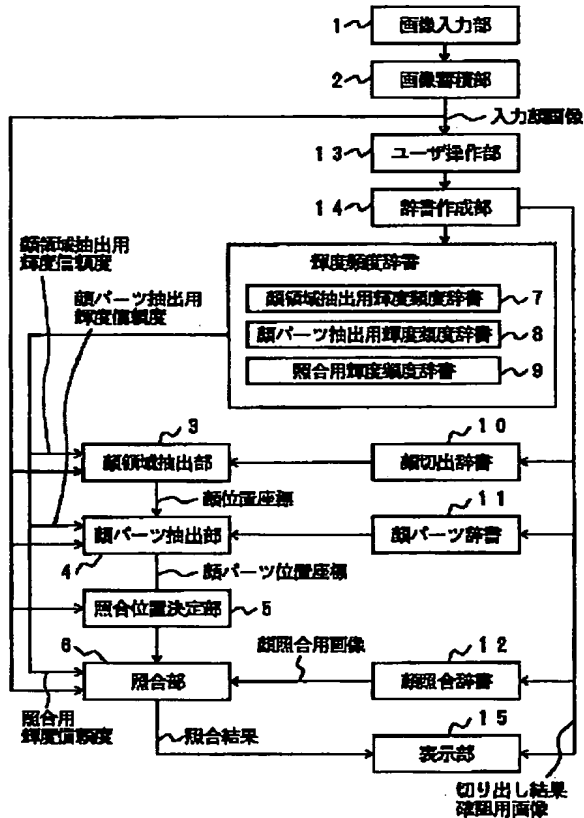
5…照合位置決定部

(14)

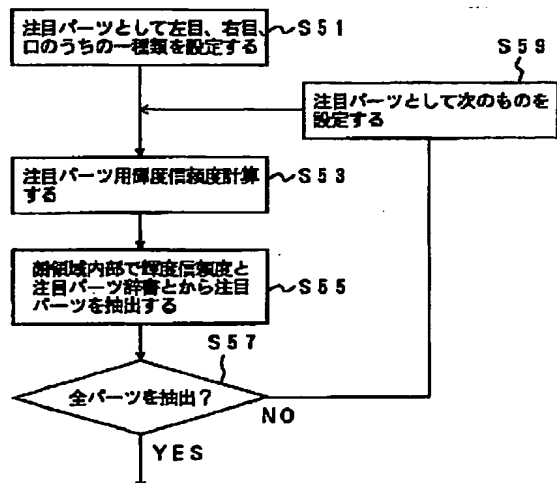
25

- 6…照合部  
 7…顔領域抽出用輝度頻度辞書  
 8…顔パーツ抽出用輝度頻度辞書  
 9…照合用輝度頻度辞書  
 10…顔切り出し辞書

【図1】



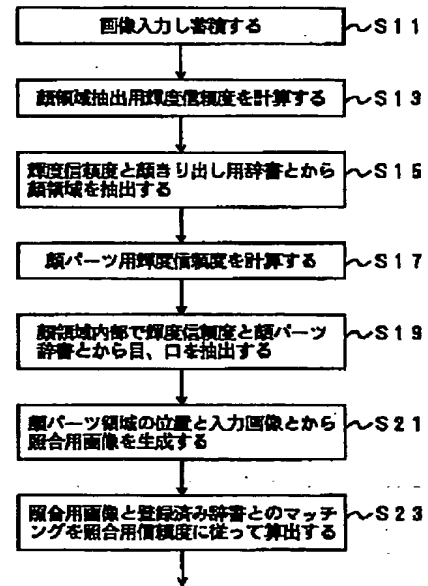
【図4】



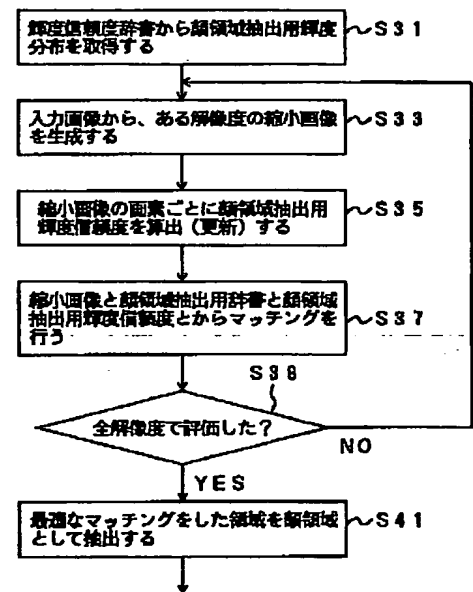
26

- 11…顔パーツ辞書  
 12…顔照合辞書  
 13…ユーザ操作部  
 14…辞書

【図2】

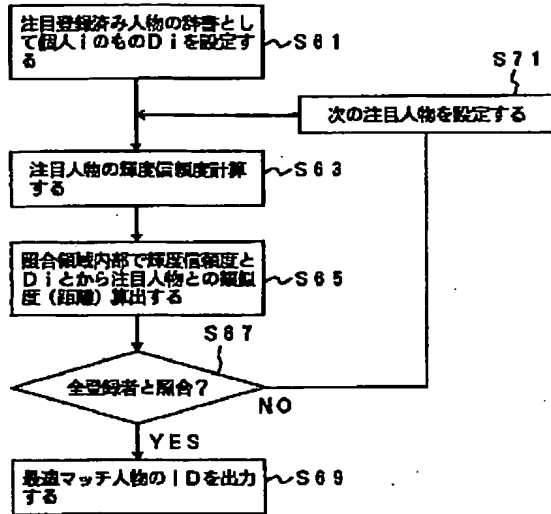


【図3】

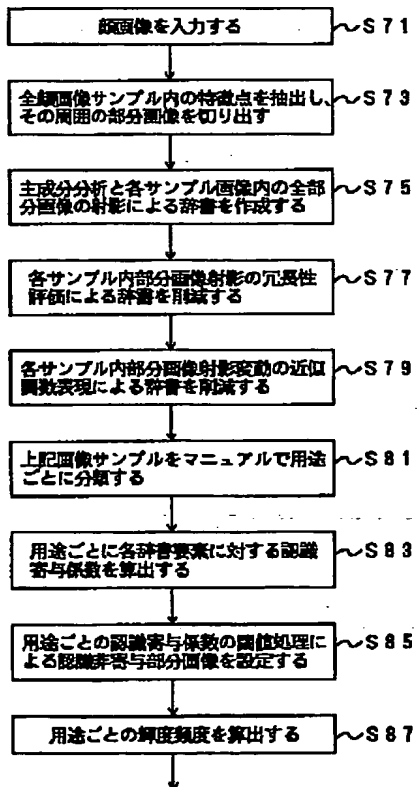


(15)

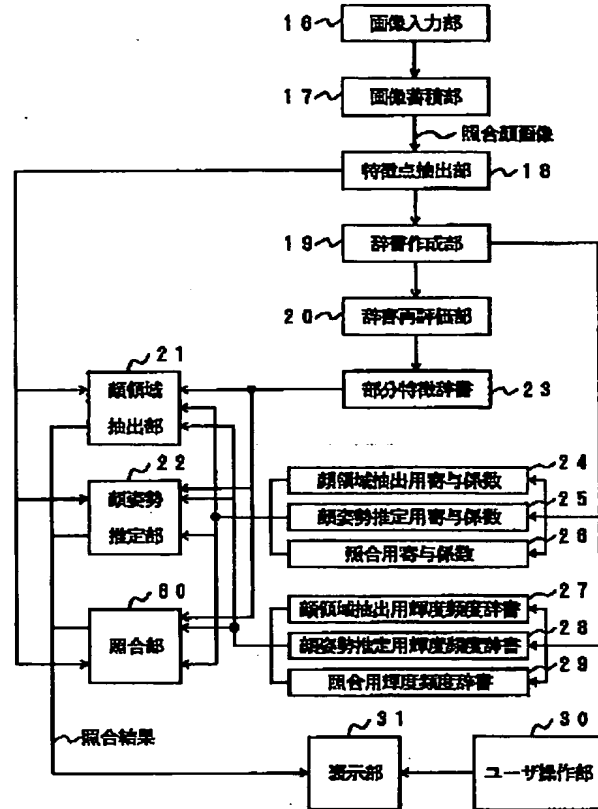
【図5】



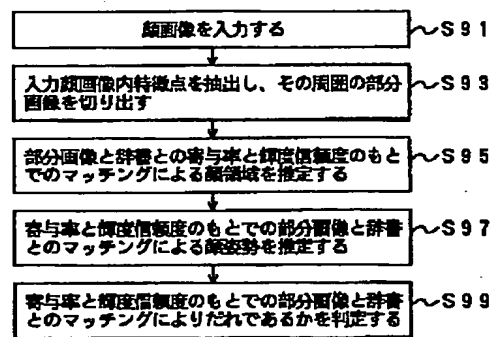
【図7】



【図6】



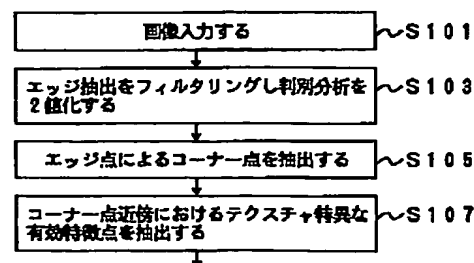
【図8】



【図10】

(a)	-1	0	1	0	-1
	0	-1	1	-1	0
	1	1	0	1	1
	0	-1	1	-1	0
(b)	-1	0	1	0	-1
	0	0	-1	0	0
	0	1	-1	0	0
	1	1	0	1	1
(c)	0	0	-1	0	0
	0	0	-1	0	0
	1	0	0	0	-1
	0	1	0	-1	0
	0	0	0	0	0
	0	-1	0	1	0
	-1	0	0	0	1

【図9】



(16)

【図11】

画像	I1	I2	I3	.....	I <sub>n-1</sub>	I <sub>n</sub>
I1	x	D12	D13	.....	D1 <sub>n-1</sub>	D1 <sub>n</sub>
I2	D21	x	D23	.....	D2 <sub>n-1</sub>	D2 <sub>n</sub>
I3	D31	D32	x	.....	D3 <sub>n-1</sub>	D3 <sub>n</sub>
.....	.....	.....	.....	x	.....	.....
I <sub>n-1</sub>	D <sub>n-1</sub> 1	D <sub>n-1</sub> 2	D <sub>n-1</sub> 3	.....	x	D <sub>n-1</sub> <sub>n</sub>
I <sub>n</sub>	D <sub>n</sub> 1	D <sub>n</sub> 2	D <sub>n</sub> 3	.....	D <sub>n</sub> <sub>n-1</sub>	x